

PCT

WORLD INTELLECTUAL PROPERTY ORGANIZATION
International Bureau

INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(51) International Patent Classification ⁵ : C07K 15/00, C12Q 1/70 G01N 33/569	A2	(11) International Publication Number: . WO 94/05700 (43) International Publication Date: 17 March 1994 (17.03.94)
(21) International Application Number: PCT/US93/08447 (22) International Filing Date: 7 September 1993 (07.09.93) (30) Priority data: 07/941,365 7 September 1992 (07.09.92) US (71) Applicant: BAYLOR COLLEGE OF MEDICINE [US/US]; One Baylor Plaza, Houston, TX 77030-3498 (US). (72) Inventors: MATSON, David, O. ; 3531 Durhill, Houston, TX 77025 (US). ESTES, Mary, K. ; 219 Carey Lane, Friendswood, TX 77546 (US). JIANG, Xi ; 9211 Kapri, Houston, TX 77025 (US). GRAHAM, David, Y. ; 4051 Mischire, Houston, TX 77054 (US).		(74) Agents: KOCH, Robert, J. et al.; Fulbright & Jaworski L.L.P., 801 Pennsylvania Avenue, N.W., Washington, DC 20004-2604 (US). (81) Designated States: AU, CA, JP, European patent (AT, BE, CH, DE, DK, ES, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE). Published <i>Without international search report and to be republished upon receipt of that report.</i>
(54) Title: METHODS AND REAGENTS TO DETECT AND CHARACTERIZE NORWALK AND RELATED VIRUSES		
(57) Abstract		
<p>Double-stranded cDNA was synthesized from nucleic acid extracted from Norwalk virus purified from stool specimens of volunteers. Single-stranded RNA probes derived from the DNA clone after subcloning into an in vitro transcription vector were also used to show that the Norwalk virus contains an ssRNA genome of about 8 kb in size. The availability of a Norwalk-specific cDNA and the genome sequence information allow rapid cloning of the entire genome and establishment of sensitive diagnostic assays. Such assays can be based on detection of Norwalk and Norwalk-related virus nucleic acids or primers and polyclonal or monoclonal antibodies to proteins expressed from the cDNA or to synthetic peptides made based on the knowledge of the genome sequence. Assays using proteins deduced from the Norwalk virus genome and produced in expression systems can measure antibody responses. Vaccines for Norwalk and related viruses are made from an expressed Norwalk virus protein.</p>		

6 TGC TCT GGG AGC GGG CAT ACA GGT TGG TGG CCA CAG GGC CTC CAA
 Cys Ser Gly Ser Gly His Thr Gly Trp Trp Arg Gln Ala Leu Gln
 21
 AGC CAA AGG TAT CAA CAA AAT TTG CAA CTG CAA GAA AAT TCT TTT
 Ser Gln Arg Tyr Gln Gln Asn Leu Gln Leu Gln Gln Asn Ser Phe
 61
 AAA CAT GAC AGG GAA ATG ATT GGC TAT CAG GTT GAA GGT TCA AAT
 Lys His Asp Arg Glu Asn Ile Gly Tyr Gln Val Glu Ala Ser Asn
 101
 CAA TTA TTG GCT AAA AAT TTG GCA ACT AGA TAT TCA CTC CTC CCG
 Gln Leu Leu Ala Lys Asn Leu Ala Thr Arg Tyr Ser Leu Leu Arg
 141
 1 GCT GGG GGT TTG ACC AGT GCT GAT GCA GCA AGA TET GTG GGA GGA
 Ala Gly Gly Leu Thr Ser Ala Asp Ala Ala Arg Ser Val Ala Gly
 181
 GCT CCA GTC ACC GGC ATT GTA GAT TGG AAT GGC GTG AGA GTG TCT
 Ala Pro Val Thr Arg Ile Val Asp Trp Asn Gly Val Arg Val Ser
 221
 GCT CCG GAG TCC TCT GCT ACC ACA TTG AGA TCC GGT GGC TTC ATG
 Ala Pro Glu Ser Ser Ala Thr Thr Leu Arg Ser Gly Gly Phe Arg
 261
 TCA GTT CCG ATA CCA TTT GGC TCT AAG CAA AAA CAG GTT CAA TCA
 Ser Val Pro Ile Pro Phe Ala Ser Lys Gln Lys Gln Val Gln Ser
 301
 TCT GGT ATT AGT AAT CCA AAT TAT TCC CCT TCA TCC ATT TCT CCA
 Ser Gly Ile Ser Asn Pro Asn Tyr Ser Pro Ser Ser Ile Ser Arg
 341
 ACC ACT AGT TGG GTC GAG TCA CAA AAC TCA TCG AGA TTT GGA AAT
 Thr Thr Ser Trp Val Glu Ser Gln Asn Ser Ser Arg Phe Gly Asn
 381
 CTT TCT CCA TAC CAC GCG GAG GCT CTC AAT ACA GTG TGG TTG ACT
 Leu Ser Pro Tyr His Ala Glu Ala Leu Asn Thr Val Trp Leu Thr
 421
 CCA GCG GGT TCA ACC
 Pro Pro Gly Asp Thr

FOR THE PURPOSES OF INFORMATION ONLY

Codes used to identify States party to the PCT on the front pages of pamphlets publishing international applications under the PCT.

AT	Austria	FR	France	MR	Mauritania
AU	Australia	GA	Gabon	MW	Malawi
BB	Barbados	GB	United Kingdom	NE	Niger
BE	Belgium	GN	Guinea	NL	Netherlands
BF	Burkina Faso	GR	Greece	NO	Norway
BG	Bulgaria	HU	Hungary	NZ	New Zealand
BJ	Benin	IE	Ireland	PL	Poland
BR	Brazil	IT	Italy	PT	Portugal
BY	Belarus	JP	Japan	RO	Romania
CA	Canada	KP	Democratic People's Republic of Korea	RU	Russian Federation
CF	Central African Republic	KR	Republic of Korea	SD	Sudan
CG	Congo	KZ	Kazakhstan	SE	Sweden
CH	Switzerland	LJ	Liechtenstein	SI	Slovenia
CI	Côte d'Ivoire	LK	Sri Lanka	SK	Slovak Republic
CM	Cameroon	LU	Luxembourg	SN	Senegal
CN	China	LV	Latvia	TD	Chad
CS	Czechoslovakia	MC	Monaco	TG	Togo
CZ	Czech Republic	MG	Madagascar	UA	Ukraine
DE	Germany	ML	Mali	US	United States of America
DK	Denmark	MN	Mongolia	UZ	Uzbekistan
ES	Spain			VN	Viet Nam
FI	Finland				

Methods and Reagents To Detect and Characterize Norwalk and Related Viruses

This application is a Continuation-in-Part of Applicant's Co-Pending U.S. Application Serial No. 07/443,492 filed November 8, 1989,
5 U.S. Application Serial No. 07/515,993, now abandoned, filed April 27, 1990, U.S. Application Serial No. 07/573,509 filed August 27, 1990, and U.S. Application Serial No. 07/696,454 filed May 6, 1991, all entitled "Methods and Reagents To Detect and Characterize Norwalk and Related Viruses."

10 This invention is supported in part through grants or awards from the Food and Drug Administration and the National Institute of Health. The United States Government may have certain rights to this invention.

Field of the Invention

The present invention relates generally to synthesizing clones of
15 Norwalk virus and calicivirus and to making probes to Norwalk and related viruses. It also relates to methods of detection and characterization of Norwalk and related viruses.

Background of the Invention

Norwalk virus is one of the most important viral pathogens causing
20 acute gastroenteritis, the second most common illness in the United States (Dingle et al., *Am. J. Hyg.* 58:16-30 (1953); Kapikian and Chanock, "Norwalk group of viruses" in B.N. Fields' 2d ed. of *Virology*, Raven Press, New York, pp. 671-693 (1990)). Up to 42% of cases of adult viral gastroenteritis have been estimated to be caused by Norwalk or
25 Norwalk-like viruses (Kaplan et al., *Ann. Internal Med.* 96(6):756-761 (1982)). Both water and foodborne transmission of Norwalk virus has been documented, and particularly large epidemic outbreaks of illness have occurred following consumption of contaminated shellfish, including clams, cockles, and oysters (Murphy et al., *Med. J. Aust.* 2:329-333 (1979);

Gunn et al., *Am. J. Epidemiol.* 115:348-351 (1982); Wilson et al., *Am. J. Public Health* 72:72-74 (1982); Gill et al., *Br. Med. J.* 287:1532-1534 (1983); DuPont *New Engl. J. Med.* 314:707-708 (1986); Morse et al., *New Engl. J. Med.* 314:678-681 (1986); Sekine et al., *Microbiol. Immunol.* 33:207-217 (1989)). An increase in fish and shellfish-related food poisonings has recently been noted and attributed to increased recognition of these entities by clinicians as well as to increased consumption of seafood (Eastaugh and Shepherd, *Arch. Intern. Med.* 149:1735-1740 (1989)).

10 Norwalk virus was discovered in 1973. Until recently, knowledge about the virus has remained limited because it has failed to grow in cell cultures and no suitable animal models have been found for virus cultivation. Human stool samples obtained from outbreaks and from human volunteer studies, therefore, are the only source of the virus. Still, 15 the concentration of the virus in stool is usually so low that virus detection with routine electron microscopy is not possible (Dolin et al., *Proc. Soc. Exp. Med. and Biol.* 140:578-583 (1972); Kapikian et al., *J. Virol.* 10:1075-1081 (1972); Thornhill et al., *J. Infect. Dis.* 132:28-34 (1975)). Current methods of Norwalk virus detection include immune 20 electron microscopy and other immunologic methods such as radio immunoassays (RIAs) or a biotin-avidin enzyme linked immunoabsorbent assays (ELISAs) which utilize acute and convalescent phase serum from humans. To date, no hyperimmune antiserum from animals has been successfully prepared due either to insufficient quantities or unusual 25 properties of the viral antigen. Preliminary biophysical characterization of virions has indicated particles contain one polypeptide (Greenberg et al., *J. Virol.* 37: 994-999 (1981)), but efforts to characterize the viral genome have failed.

Viruses related to Norwalk virus include small round enteric 30 viruses, such as viruses with typical calicivirus morphology and the astroviruses. The classification scheme for the human small enteric viruses shown in Table 1 here is an updated version of a scheme outlined

by Caul and Appleton in the *Journal of Medical Virology*, 9:257-265 (1982). This system is referred to in Cubitt *et al.*, *J. Infectious Diseases*, 156:806-814 (1987); Table 1 of the article by Appleton entitled "Small round viruses: classification and role in food-borne infections", in the book

5 Novel Diarrhoea Viruses, Ciba Foundation Symposium No. 128, pp. 108-125 (John Wiley & Sons, N.Y. (1987)); and Table 1 of the chapter entitled "Norwalk group of viruses" by Kapikian and Chanock from the book Virology (B.N.Fields, 2d ed., Raven Press (1990)).

As shown in Table 1, human small round structured enteric viruses

10 include calicivirus and astrovirus. The recent sequencing of Norwalk virus indicates that Norwalk virus is a calicivirus and has a genome organization like that of other caliciviruses. In addition to the human small round enteric viruses are a large number of non-human small round

15 round structured viruses based upon their morphology. Examples of these viruses are the primate calicivirus isolated from the pygmy chimpanzee, described in the journal *Science* 221:79-81 (1983), a porcine enteric calicivirus, described in the *Journal of Clinical Microbiology* 12:105-111 (1980), and bovine astroviruses described in *Vet Pathol.* 21:208-215 (1984).

20 Individual calicivirus types will at times exhibit host specificity and tissue tropisms, but as an overall group they cause gastroenteritis, hepatitis, abortion, skin lesions, pneumonia, myocarditis, and encephalitis. The caliciviruses infecting humans fit in this context in that Norwalk-like viruses cause gastroenteritis, hepatitis E causes hepatitis, and San Miguel

25 sea lion virus type 5 causes skin vesicles in humans as well as infections in seals, fish, pigs and cattle. (D. O. Matson "Calicivirus Infections" in *Textbook of Pediatric Infectious Disease*, 3d ed., R. D. Feigin and J. D. Cherry, eds., W. B. Sanders, Philadelphia, (in press)).

Summary of the Invention

It is therefore an object of the invention to detect and characterize the Norwalk and related virus genomes by synthesizing and cloning a cDNA library.

5 It is an associated object of the invention to deduce amino acid sequences from Norwalk and related viral cDNA.

Another object of the invention is to develop probes or primers to confirm the genetic relationship between the Norwalk virus and the Norwalk-related viruses.

10 Still another object of the invention is to develop a method of preparing polyclonal and monoclonal antibodies to the Norwalk and related viruses.

Yet still another object of the invention is to develop a method of making probes to detect Norwalk and related viruses.

15 A further object of the invention is to use the cDNA or fragments or derivatives thereof in assays to detect Norwalk and related viruses in samples suspected of containing the viruses.

A still further object of the invention is to express proteins to measure antibody responses.

20 A nucleotide sequence of the genome sense strand of the Norwalk virus cDNA clone intended to accomplish the foregoing objects includes the nucleotide sequence shown in Table 2. Within the Norwalk nucleotide sequence are regions which encode proteins. The nucleotide sequence of the Norwalk virus genome, its fragments and derivatives are used to make
25 diagnostic products, vaccines and antivirals.

Other and still further objects, features and advantages of the present invention will be apparent from the following description of a presently preferred embodiment of the invention.

Brief Description of the Figures

30 Figure 1. EM picture of Norwalk and related viruses. Norwalk virus (A),

human Calicivirus (B), small round structured virus (C), and human astrovirus (D). The var is 0.1 μ m.

Figure 2a. Hybridization of stool samples with 32 P-labeled plasmid DNA for screening positive Norwalk cDNA clones. Nucleic acids from paired stools [before (b) and after (a) infection with Norwalk virus] from two volunteers (1 and 2) were dotted on Zetabind filters. Replicate strips were prepared and hybridized at 50°C and 65°C with each test clone (pUC-27, pUC-593, pUC-13 and pUCNV-953). One clone (pUCNV-953) which reacted only with stool samples after (but not before) Norwalk infection was considered as a potential positive clone and was chosen for further characterization.

Figure 2b. Dot blot hybridization of clone 32 P-labeled pUCNV-953 with another 3 sets of stool samples collected at different times after infection (B = before acute phase of illness; A = acute phase of illness; P = post-acute phase of illness) of 3 volunteers. The nucleic acids were dotted directly or after treatment with RNase or with DNase before dotting. Double-stranded homologous cDNA (pUCNV-953) was dotted after the same treatments as the stool samples.

Figure 3. Dot blot hybridization of Norwalk viruses in a CsCl gradient with ssRNA probes made from pGEMNV-953. Aliquots of 50ul from each fraction in a CsCl gradient were dotted onto a Zetabind filter. Duplicates of filters were made and hybridized with the two ssRNA probes respectively. The two strands were subsequently called cRNA (positive hybridization with the viral nucleic acid) and vRNA (no hybridization with the viral nucleic acid, data not shown). The graph shows EM counts of Norwalk viruses from each fraction of the same CsCl gradient for the dot blot hybridization. Five squares from each grid were counted and the average of the number of viral particles per square was calculated.

Figure 4. The nucleotide sequence of the genome sense strand of the first Norwalk virus cDNA clone. The deduced amino acid sequence of a long open reading frame in this cDNA also is shown.

Figure 5. Schematic diagram of Norwalk cDNA clones. pUCNV-953 was
5 the first positive clone identified. Overlapping clones were determined by restriction enzyme analyses and partial sequencing of the clones. AAA indicates the poly(a) tail at the 3' end of the viral genome.

Figure 6. Norwalk virus encodes an RNA-directed RNA polymerase
sequence motif. The deduced amino acid sequence of a portion of Norwalk
10 virus pUCNV-4095 (NV) is compared with consensus amino acid residues thought to encode putative RNA-directed RNA polymerases of hepatitis E virus (HEV), hepatitis C virus (HCV), hepatitis A virus (HAV), Japanese encephalitis virus (JE), poliovirus (polio), foot-and-mouth disease virus (FMD), encephalomyocarditis virus (EMC), Sindbis virus (SNBV), tobacco
15 mosaic virus (TMV), alfalfa mosaic virus (AMV), brome mosaic virus (BMV), and cowpea mosaic virus (CpMV). Sequences for viruses other than NV are from Figure 3 of Reyes *et al.*, Science 247:1335-1339 (1990).

Figure 7. Three pairs of initial primers used to amplify the Norwalk virus
genome. RNA was extracted from a stool sample (sample 543-11) by the
20 CTAB technique and amplified by RT-PCR. Lanes 1 and 5, 1-kb markers from Bethesda Research Laboratories (the markers that migrated as 1.6, 1.0 and 0.5 kb are labeled); lane 2, PCR with Norwalk virus primers 8 and 9; lane 3, PCR with Norwalk primers 16 and 17; lane 4, PCR with Norwalk primers 1 and 4. The amplified products were separated on the
25 agarose gel and visualized with UV light after staining with ethidium bromide. The small product seen in lane 3 was made in variable amount in different experiments. The positions of the three primer pairs used in this study are given above the autoradiograph. The numbers below the map indicate the size (in base pairs) of the RT-PCR product.

Figure 8. This schematic shows the organization of Norwalk genome given in Table 2. The features shown here are based on analyses of the nucleotide sequence of the Norwalk virus genome and the deduced amino acid sequence of proteins encoded in the genome. The genome contains 7753 nucleotides including 111 A's at the 3'-end. Translation of the sequence predicts that the genome encodes three open reading frames (shown by the open boxes in the second line). The first open reading frame is predicted to start from an initiation codon at nucleotide 146 and it extends to nucleotide 5359 (excluding the termination codon). The second open reading frame is initiated at nucleotide 5346 and it extends to nucleotide 6935, and a third open reading frame exists between nucleotides 6938 and 7573. Based on comparisons of these predicted proteins with other proteins in the protein databank, the first open reading frame is a protein that is eventually cleaved to make at least three proteins. These three proteins include a picornavirus 2C-like protein, a 3C-like protease and a 3D-like RNA-dependent RNA polymerase. The second open reading frame encodes the capsid protein, which contains sequence homology with the picornavirus VP3 protein.

Figure 9. Nucleotide and amino acid sequence of human calicivirus Sapporo cDNAs. The 551 nucleotide known sequence of human calicivirus Sapporo (HuCV Sapporo) is presented in its entirety. Below the nucleotide sequence is the amino acid sequence for HuCV Sapporo. Above the HuCV Sapporo nucleotide sequence is the sequence of the cDNA from a Houston day care center outbreak (Day care). In the Day care sequence a "." indicates the nucleotide is identical to the HuCV Sapporo nucleotide at that site. Where a nucleotide difference occurred in the Day care sequence, a new letter is indicated at that position. "N" indicates uncertainty of the nucleotide at that site. Below the HuCV Sapporo amino acid sequence are arrows indicating the extent of cDNAs at 23s2m31 and c-29_4-gel (which together contribute to the 551 nucleotides of the known sequence) and the new 36 primer (see Table 6).

Figure 10. Nucleotide homologies between calicivirus cDNAs and calicivirus strains with known sequences. All comparisons are in reference to the sequence of human calicivirus Sapporo. The length of the baseline indicates the known sequence region. The boxes indicate areas of nucleotide sequence homology between HuCV Sapporo and the indicated strain. The length of the box indicates the part of the indicated strain where homology exists and the height of the box indicates the strength of the homology. SD = standard deviation. SD 3 or greater is significant. The numbers under the Norwalk homology box indicate the region of the Norwalk virus genome where homology was observed.

Figure 11. Strategy used to obtain nucleotide sequence of the Norwalk-related virus SRSV/KY/89 using primers from the Norwalk virus sequence. This figure shows a partial schematic of the Norwalk virus genome and the predicted ORF1 showing the location of the 3D-like polymerase region, the second ORF showing the location of the VP3-like domain and the start of ORF 3. On the bottom, the solid lines show regions of KY89 sequenced based on using primer sets (see numbers such as 36 and 35, etc) chosen from the sequence of the Norwalk virus genome.

Figure 12. Comparison of the Norwalk virus nucleotide sequence with the Norwalk virus-related virus SRSV/KY/89 nucleotide sequence. Part of the nucleotide sequence of Norwalk-related virus SRSV/KY/89 was determined using primers from the Norwalk-virus (NV) genome. Primers from the NV genome used to obtain the sequence of this Norwalk-related virus are shown in Table 6. Some of these primers were modified based on the initial nucleotide sequence obtained from the SRSV/KY/89 to obtain the rest of the sequence of SRSV/KY/89. The primers shown here and in Table 6 are used by way of example only; other NV primers can be used.

Figure 13. Comparison of deduced amino acid sequence of proteins of the Norwalk virus and the Norwalk-related virus SRSV/KY/89. The protein

sequence of SRSV/KY/89 was deduced from the nucleotide sequence shown in Figure 12. Figure 13a shows a comparison of the deduced amino acid sequence of ORF2, the capsid, of SRSV/KY/89 with the same region encoded in the Norwalk virus genome. Figure 13b shows a comparison of the deduced amino acid sequence of part of the polymerase protein of SRSV/KY/89 with that of Norwalk virus. Comparisons of similar sequences from other Norwalk-related viruses will permit discovery of conserved and divergent regions including antigenic regions. The information will rapidly permit choices of broadly reactive primers to detect all Norwalk-related viruses and specific primer sets to detect individual Norwalk-related viruses. Similarly, fragments and peptides with common amino acid sequences or specific amino acid sequences can be selected for development of diagnostics, vaccines and antivirals.

Figure 14. Comparison of partial nucleotide sequences of Norwalk virus and six Norwalk-related viruses obtained using primers from the NV genome. Sequences from SRSV/CDC 6/91, SRSV/UT/88, SMA/78; SRSV/Cambridge, UK/92, SRSV/CDC 32, Norwalk virus/68, SRSV-3/88, SRSV/KY89/89. Figures 14a and 14b show two different regions of the genome.

Figure 15. Expression of the Norwalk virus capsid protein. Baculovirus recombinants (C-6 and C-8) that contain a subgenomic piece of Norwalk virus DNA (from nucleotides 5337 to 7753) were selected and used to infect insect (*Spodoptera fugiperda*) cells at a multiplicity of infection of 10 PFU/cell. After 4 days of incubation at 27°C, the infected cells were harvested and the proteins were analyzed by electrophoresis on 12% polyacrylamide gels. The proteins were visualized after staining with Coomassie blue. The Norwalk-expressed protein (highlighted by the arrowhead) is only seen in the recombinant-infected cells, but not in wild-type baculovirus (wt) or mock-infected (m) insect cells.

Figure 16. The Norwalk virus expressed protein shows immunoreactivity with sera from volunteers infected with Norwalk virus. The expressed protein shown in Figure 11 was absorbed onto the wells of a 96-well ELISA plate and its reactivity was tested with dilutions of serum samples taken from volunteers before (pre) and three weeks after (post) infection with Norwalk virus. After an incubation at 37°C for 2 hours, a peroxidase-conjugated goat-anti-human IgG, IgM and IgA serum was added and reactivity was subsequently observed by reading the optical density at 414nm after addition of the substrate. The data show that post-infection sera reacted strongly with the expressed antigen at serum dilutions of 1:100 and 1:1000, and some sera were still specifically reactive at a dilution of 1:10,000.

Figure 17. Baculovirus recombinants containing the 3'-end of the Norwalk genome produce virus-like particles in insect cells. Lysates from insect cells infected with baculovirus recombinant C-8 (see Figure 11) were analyzed by electron microscopy and shown to contain numerous virus-like particles. These particles are the same size as virus particles obtained from the stools of volunteers infected with Norwalk virus. Bar = 50 nm.

Figure 18. Norwalk virus-like particles can be purified in gradients of CsCl. Supernatants of insect cells infected with the baculovirus recombinant C-8 were processed by extraction with genetron and PEG precipitation and virus eluted from these PEG pellets was centrifuged in CsCl gradient in a SW50.1 rotor for 24 hours at 4°C. The gradient was fractionated and material in each fraction was adsorbed onto two wells of an ELISA plate. Duplicate wells were then treated either with pre- or post-infection serum, peroxidase-conjugated goat anti-human serum and substrate and the reactions were monitored by reading the OD_{414nm}. A peak was observed in the gradient at a density of 1.31 g/cm³ and this peak was shown to contain virus-like particles by electron microscopy. This peak also contained a major protein of an approximate molecular weight

of 58,500 that co-migrated with the protein expressed in the insect cells from the same baculovirus recombinant.

Figure 19. Use of the expressed virus-like particles to measure the reactivity of pre- and post-serum samples from volunteers infected with Norwalk virus shows that most volunteers have an immune response. Volunteer 6 who did not show an immune response also did not become ill after being administered virus.

Figure 20. Partial sequence of the primate *Pan paniscus* cDNA atprcvw2.

Detailed Description of the Invention

10 It is readily apparent to one skilled in the art that various substitutions and modifications may be made to the invention disclosed herein without departing from the scope and spirit of the invention.

The term "fragment" as used herein is defined as any portion of the Norwalk virus genome or a subgenomic clone of the Norwalk virus that
15 is required to be expressed to produce or encodes a peptide which in turn is able to induce a polyclonal or monoclonal antibody. It is possible a peptide of only 5 amino acids could be immunogenic but usually peptides of 15 amino acids or longer are required. This depends on the properties of the peptide and it cannot be predicted in advance.

20 The term "derivative" as used herein is defined as larger pieces of DNA or an additional cDNA which represents the Norwalk virus genome and which is detected by direct or sequential use of the original cDNA and any deduced amino acid sequences thereof. Clone pUCNV-1011, therefore, is a derivative, although it does not overlap or share sequences with the
25 original clone. Also included within the definition of derivative are RNA counterparts of DNA fragments and DNA or cDNA fragments in which one or more bases have been substituted or to which labels and end structures have been added without affecting the reading or expression of the DNA or cDNA.

The terms Norwalk "related viruses" and "Norwalk-like viruses" as used herein are defined as human and non-human calicivirus, astrovirus and small round structured viruses (SRSV). As the genomic sequences of most of these viruses are not known, this classification is based on morphology as described by Caul and Appleton in the *Journal of Medical Virology*, 9:257-265 (1982); by Appleton in the article entitled "Small round viruses: classification and role in food-borne infections", in the book Novel Diarrhoea Viruses, Ciba Foundation Symposium No. 128, pp. 108-125 (John Wiley & Sons, N.Y. (1987)); and by Kapikian and Chanock in the chapter entitled "Norwalk group of viruses" from the book Virology (B.N.Fields, 2d ed., Raven Press (1990)). As the genomic sequences of the viruses become known, those skilled in the art will be able to determine Norwalk-related viruses and Norwalk-like viruses based on nucleotide homologies.

Within the Norwalk-related viruses is a subgroup of viruses referred to herein as the SRSV's or the Norwalk group. The Norwalk group includes Snow Mountain Agent (SMA), Hawaii Agent, Taunton Agent, Amulree, Otofuke, and Montgomery County Agent. The Norwalk group is characterized by small, round, structured viruses with an amorphous surface or ragged outline.

Production of Norwalk Virus for Molecular Cloning

Norwalk virus was produced by administration of safety tested Norwalk virus (8FIIa) to adult volunteers. The virus inoculum used in the volunteer study, was kindly supplied by Dr. Albert Kapikian (Laboratory of Allergy and Infectious Diseases, National Institutes of Health, Bethesda, MD). This virus originated from an outbreak of acute gastroenteritis in Norwalk, Ohio (Dolin et al., 1971). Two ml of a 1 to 100 dilution of 8FIIa in TBS was administered orally to each individual with 80 ml of milli-Q water (Millipore, Bedford, MA 01730). Sodium bicarbonate solution was taken by each person 2 minutes before and 5 minutes after virus administration. The volunteer studies were approved by the Institutional

Review Board for Human Research at Baylor College of Medicine, at the Methodist Hospital and at the General Clinical Research Center. The virus was administered to the volunteers in the General Clinical Research Center where the volunteers were hospitalized and under extensive
5 medical care for 4 days. All stools were collected and kept at -70°C for later use.

Purification of Norwalk Viruses from Stool Samples

A 10% solution of stool samples in TBS was clarified by low speed centrifugation at 3000 rpm for 15 minutes. The resulting supernate then
10 was extracted two to three times with genetron in the presence of 0.5% Zwittergent 3-14 detergent (Calbiochem Corp., La Jolla, CA). Viruses in the aqueous phase were concentrated by pelleting at 36,000 rpm for 90 minutes through a 40% sucrose cushion in a 50.2 Ti rotor (Beckman Instruments, Inc., Palo Alto, CA 94304). The pellets were suspended in
15 TBS and mixed with CsCl solution (refractive index 1.368) and centrifuged at about 35,000 rpm for about 24 hours in a SW50.1 rotor (Beckman). The CsCl gradient was fractionated by bottom puncture and each fraction was monitored for virus by EM examination. The peak fractions containing Norwalk virus were pooled and CsCl in the samples was
20 diluted with TBS and removed by pelleting the viruses at about 35,000 rpm for 1 hour. The purified virus was stored at about -70°C.

Extraction of Nucleic Acids from Purified Virus

One method of extraction involved treating purified Norwalk virus from CsCl gradients with proteinase K (400 ug/ml) in proteinase K buffer
25 (0.1 M Tris-Cl pH 7.5, 12.5 mM EDTA, 0.15 M NaCl, 1% w/v SDS) at about 37°C for about 30 minutes. The samples were then extracted once with phenol-chloroform and once with chloroform. Nucleic acids in the aqueous phase were concentrated by precipitation with 2.5 volumes of ethanol in the presence of 0.2 M NaOAc followed by pelleting for 15
30 minutes in a microcentrifuge.

cDNA Synthesis and Cloning of Amplified of cDNA

One method of synthesis and cloning included denaturing nucleic acids extracted from the purified Norwalk viruses with 10 mM CH_3HgOH . Then cDNA was synthesized using the cDNA synthesis kit with the
5 supplied random hexanucleotide primer (Amersham, Arlington Heights, IL 60005). After the second strand synthesis, the reaction mixture was extracted once with phenol-chloroform and once with chloroform followed by ethanol precipitation. Amplification of DNA was performed using the random prime kit for DNA labeling (Promega Corp., Madison, WI
10 53711-5305). Eight cycles of denaturation (100°C for 2 minutes), reannealing (2 minutes cooling to room temperature) and elongation (room temperature for 30 minutes) were performed after addition of Klenow fragment (Promega Corp.). A DNA library was constructed in pUC-13 with blunt-end ligation into the Sma I site.

15 Screening of the Library for Positive Clones

As one method of screening, white colonies from transformed DH5 alpha bacterial cells (BRL) were picked and both a master plate and minipreps of plasmid DNA were prepared for each clone. Clones containing inserts were identified after electrophoresis of the plasmid
20 DNA in an agarose gel. The insert DNA in the agarose gel was cut out and labeled with ^{32}P using random primers and Klenow DNA polymerase such as in the PRIME-A-GENE® labeling system (Promega Corp.). Other isotopic or biochemical labels, such as enzymes, and fluorescent, chemiluminescent or bioluminescent substrates can also be used. Nucleic
25 acids extracted from paired stool samples (before and after Norwalk infection) from two volunteers (543 and 544) were dotted onto Zetabind filters (AFM, Cuno, Meriden, CT). Replicate filter strips were prepared and hybridized with each labeled plasmid probe individually at 65°C without formamide. Potential positive clones were judged by their
30 different reactions with the pre- and post-infection stools. Clones which reacted with post (but not pre-) infection stools of volunteers were considered positive and these clones on the master plates were

characterized further. Once one Norwalk clone was identified, it was used to rescreen the cDNA library to identify additional overlapping clones. Rescreening the cDNA library with these additional clones can ultimately identify clones representing the entire Norwalk virus genome.

5 Reverse Transcriptase-Polymerase Chain Reaction Production of cDNA Clones from Viruses Related to Norwalk Virus

One method for producing cDNA clones of viruses related to Norwalk virus using the knowledge of the Norwalk virus genome sequence is the reverse transcription-polymerase chain reaction method. In this
10 procedure, RNA was extracted from 300 uL of specimen containing the related virus. Complementary DNA was prepared by reverse transcriptase-polymerase chain reaction (RT-PCR) using a primer pair (for example primers 36 and 35 shown in Table 6) derived from the sequence of Norwalk virus. The resulting product was ligated into a plasmid vector
15 and transfected into *E. coli*. Plasmids then were partially purified from the bacteria and the inserted PCR product was sequenced in the plasmid by dideoxy chain termination to examine the relation to Norwalk virus by nucleotide and predicted protein homology.

The following examples are offered by way of illustration and are not
20 intended to limit the invention in any manner.

Example 1

Electron micrograph confirmation

To permit better diagnosis and molecular characterization of Norwalk virus and related viruses, a cDNA library for Norwalk was
25 derived from nucleic acid extracted from virions purified from stool samples. Norwalk virus was purified with methods used previously for hepatitis A and rotaviruses from stool samples with some modifications (Jiang et al., 1986). Basically, stool samples obtained from volunteers administered Norwalk virus were treated with genetron to remove lipid
30 and water insoluble materials. Virus in the aqueous phase was then pelleted through a 40% sucrose cushion. The resulting pellets were

resuspended, sonicated and loaded in a CsCl gradient for isopycnic centrifugation.

Figure 1 shows an electron micrographs of purified Norwalk viruses isolated by the above procedure and Norwalk-related viruses used to
5 produce cDNAs using RT-PCR.

Example 2

Initial cDNA synthesis, cloning and screening

A cDNA library was generated from nucleic acids extracted from these purified viruses by proteinase K treatment of the samples followed
10 by phenol-chloroform extraction and ethanol precipitation (Jiang et al., 1986; 1987). Because the nature of the viral genome was unknown, the extracted nucleic acids were denatured with methylmercuric hydroxide before cDNA synthesis. Random primed cDNA was synthesized with the Gubler-Hoffman method (cDNA synthesis system plus, Amersham) and a
15 small amount of cDNA was obtained. Direct cloning of this small amount of cDNA was unsuccessful. Therefore, a step of amplification of the DNA was performed by synthesizing more copies of the DNA with random primers and the Klenow fragment of DNA polymerase before cloning. The procedure involved cycles of denaturation, addition of random primers and
20 the Klenow fragment of DNA polymerase, reannealing and elongation. With this procedure, a linear incorporation of labeled nucleotides into product was observed as the number of cycles of synthesis was increased. The number of cycles performed was limited (<10) to avoid the synthesis of an excess of smaller fragments. In the case of Norwalk cDNA, eight
25 cycles of amplification were performed and approximately 2.5 ug of DNA were obtained, which was at least a 100-fold amplification of the starting template cDNA. This amplified cDNA was cloned into pUC-13 by blunt-end ligation and a positive clone (pUCNV-953) was isolated.

To obtain the positive Norwalk virus clone, minipreparations of the
30 plasmid DNAs containing potential inserts were screened by agarose gel electrophoresis. Inserts of the larger clones in the gel were cut out and probes were made with the DNA in the gel using the PRIME-A-GENE®

labeling system (Promega Corp.). These probes were hybridized individually with paired stool samples (before and after Norwalk infection) from two volunteers (Figure 2a). One clone (pUCNV-953) reacted with post- but not pre-infection stool samples from both volunteers.

5

Example 3

Confirmation of viral origin of the clone pUCNV-953

To further confirm the viral origin of the clone pUCNV-953, six more paired stool samples were tested and the same results were obtained. Figure 2b shows a dot blot hybridization of the clone with stool samples
10 collected at different times post-infection of the disease. Strong signals were observed only with stools from acute phase, but not before and after the illness. This result was consistent with previous RIA assays for viral antigen detection using convalescent sera from volunteers with Norwalk diarrhea and immune electron microscopy (IEM) studies of the samples
15 for viral particle examination. This result also agrees with the patterns of virus shedding in stool in the course of the disease (Thornhill et al., 1975). When the pUCNV-953 clone was hybridized with fractions of a CsCl gradient from the Norwalk virus purification scheme, an excellent correlation between hybridization and EM viral particle counts was
20 observed (Figure 3). The peaks of the hybridization signals and viral particle counts both were at fractions with a density of 1.38 g/cm^3 , which agrees with previous reports of the biophysical properties of Norwalk virus. Finally, the clone was tested by hybridization with highly purified Norwalk virus electrophoresed on an agarose gel. A single hybridization
25 band was observed with Norwalk virus but not with HAV and rotavirus. Sequence analysis of the pUCNV-953 cDNA showed this clone is 511 bp (Figure 4). This partial genomic cDNA encodes a potential open reading frame for which the amino acid sequence has been deduced (Figure 4). No significant nucleotide or deduced amino acid sequence homology was
30 found by comparison with other sequences in the Gen Bank (Molecular Biology Information Resource, Eugene Software, Baylor College of Medicine).

Example 4

Use of Norwalk virus cDNA to characterize the viral genome

The pUCNV-953 cDNA was subcloned into the transcription vector pGEM-3Zf(+) and grown. ssRNA probes were then generated by in vitro
5 transcription using SP6 and T7 polymerases (Bethesda Research Laboratory). When two opposite sense ssRNA probes were hybridized with the viral nucleic acid separately, only one strand reacted with the virus, indicating the viral genome is single-stranded. As shown in Figure 2b, the hybridization signals were removed by treatment of the viral
10 nucleic acid with RNase (but not with DNase) before loading them onto the filters, indicating the virus genome contains ssRNA. A long open reading frame was found in one of the two strands of the inserted DNA by the computer analysis of the sequences of pUCNV-953. The ssRNA probe with the same sequence as this coding strand does not react with
15 the viral nucleic acid, but the complementary ssRNA probe does react in the hybridization tests. Therefore, Norwalk virus contains a positive sense single-stranded RNA genome. The size of the genome of Norwalk virus was estimated to be about 8 kb based on comparisons of the migration rate of the purified viral RNA in agarose gels with molecular
20 weight markers.

The pUCNV-953 cDNA was used to rescreen a second cDNA library made as follows. A clone of the Norwalk or related virus was synthesized by isolating nucleic acid from purified Norwalk virus; cDNA was synthesized using reverse transcriptase and random primers; a second
25 strand of DNA was synthesized from the cDNA; and at least one copy of DNA was inserted into a plasmid or a cloning and expression vector; and screening the library with the original puCNV-953 cDNA identified clones containing fragments of (or the complete) Norwalk or related genome. Alternatively at least one copy of DNA was inserted in a cloning and
30 expression vector, such as lambda ZAPII® (Stratagene Inc.), and the cDNA library was screened to identify recombinant phage containing fragments of or the complete Norwalk or related genome. Additional cDNAs were made and found with this method. Use of these additional cDNAs to

made and found with this method. Use of these additional cDNAs to rescreen the library resulted in detection of new clones (Figure 5).

Thus, those skilled in the art will recognize that entire Norwalk virus cDNA sequence, or fragments or derivatives thereof, can be used in
5 assays to detect the genome of Norwalk and other related viruses. The detection assays include labeled cDNA or ssRNA probes for direct detection of the Norwalk virus genome and measurement of the amount of probe binding. Alternatively, primers or small oligonucleotide probes (10 nucleotides or greater) and polymerase chain reaction amplification
10 are used to detect the Norwalk and Norwalk-related virus genomes. Expression of the open reading frame in the cDNA is used to make hyperimmune or monoclonal antibodies for use in diagnostic products, vaccines and antivirals.

Using the above methodology, the nucleotide sequence in Table 2
15 was identified. Within that nucleotide sequence, the encoding regions for several proteins have been identified. In that sequence, the first protein is encoded by nucleotides 146 through 5339 and the amino acid sequence is shown in Table 3. This first protein is eventually cleaved to make at least three proteins including a picornavirus 2C-like protein, a 3C-like
20 protease and an RNA-dependent RNA polymerase. The RNA-dependent RNA polymerase is deduced from nucleotides 4543 to 4924 of the Norwalk virus genome as shown in Table 3. The fact that this portion of the genome contains an RNA polymerase is verified by comparisons with RNA polymerase in other positive sense RNA viruses (Figure 6 SEQ ID NOS
25 38 through 50).

Also in the sequence in Table 2, two other protein encoding regions were found. They are encoded by nucleotides 5346 through 6935 and nucleotides 6938 through 7573. The amino acid sequences for these two proteins are shown in Tables 4 and 5, respectively.

Example 5

Diagnostic assays based on detection of the
sequences of the Norwalk virus genome

Hybridization assays are the assays of choice to detect Norwalk virus
5 because small amounts of virus are present in clinical or contaminated
water and food specimens. Previously, detection of Norwalk and related
nucleic acids was not possible because the genome of Norwalk virus was
not known and no sequence information was available. Probes made from
the Norwalk virus cDNA or primers made from the Norwalk virus genome
10 sequence allow methods to amplify the genome for diagnostic products to
be established. Probes to identify Norwalk virus alone and to identify
other Norwalk-related viruses enable development of either specific assays
for Norwalk or general assays to detect sequences common to many or all
of the Norwalk-related agents.

15 In the past, one major difficulty encountered in RT-PCR detection
of viral RNA in stool samples was that uncharacterized factor(s) are
present in stools which inhibit the enzymatic activity of both reverse
transcriptase and Taq polymerase (Wilde et al., *J. Clin. Microbiol.*
28:1300-1307, 1990). These factor(s) were difficult to remove by routine
20 methods of nucleic acid extraction. Techniques were developed using
cetyltrimethylammonium bromide (CTAB) and oligo d(T) cellulose
specifically to separate viral RNA from the inhibitory factor(s). These
techniques were based on the unique properties of CTAB which selectively
precipitates nucleic acid while leaving acid insoluble polysaccharide in the
25 supernatant. The resulting nucleic acid was further purified by adsorption
onto and elution from oligo d(T) cellulose. This step removes unrelated
nucleic acids that lack a poly(A) tail. With this technique, Norwalk virus
was detected easily by PCR in very small amounts (400 ul of a 10%
suspension) of stool sample. For example, one skilled in the art will
30 recognize that it is now possible to clone the genome of RNA viruses
present in low concentrations in small amounts of stool after RT-PCR and
a step of amplification of the viral RNA by RT-PCR using random
primers. In some cases, RT-PCR active nucleic acids are extracted with

CTAB and without oligo d(T) cellulose. In addition, now that the inhibitor(s) can be removed from stool, it is also possible to detect and clone nucleic acids of other viruses (DNA viruses, non-poly(A) tailed RNA viruses) present in stool.

5 The CTAB and oligo d(T) cellulose technique of extraction followed by detection of viral RNA with RT-PCR was used on stool samples and could be used on water and food samples. Stool sample was suspended in distilled water (about 10% wt/vol) and extracted once with genetron. Viruses in the supernatant were precipitated with polyethylene glycol at
10 a final concentration of about 8%. The viral pellets were treated with proteinase K (about 400 ug/ml) in the presence of SDS at about 37°C for about 30 minutes followed by one extraction with phenol chloroform and one with chloroform. A solution of about 5% CTAB and about 0.4M NaCl was added at a ratio of sample:CTAB equal to about 5:2. After incubation
15 at about room temperature for about 15 minutes and at about 45°C for about 5 minutes, the nucleic acids (including the viral RNA) were collected by centrifugation in a microcentrifuge for about 30 minutes. The resultant pellets were suspended in about 1M NaCl and extracted twice with chloroform. The viral RNA in the aqueous phase was used directly
20 in RT-PCR reactions or further purified by adsorption/elution on oligo d(T) cellulose.

A batch method of adsorption/elution on oligo d(T) cellulose was used to purify poly(A) tailed RNA. In this procedure, nucleic acids partially purified as described above or RNA extracted directly with
25 phenol chloroform (without CTAB treatment) were mixed with oligo d(T) cellulose (about 2-4mg/sample) in a binding buffer (about 0.5M NaCl and 10mM Tris, pH 7.5). The mixture was incubated at about 4°C for about 1 hr with gentle shaking and then centrifuged for about 2 minutes in a microcentrifuge. The oligo d(T) cellulose pellet was washed 3-4 times with
30 binding buffer and then the poly(A) tailed RNA was eluted with 1X TE buffer (about 10mM Tris, 1mM EDTA, pH 7.5). The supernate was collected following centrifugation to remove the oligo d(T) cellulose and the viral RNA in the supernate was precipitated with ethanol. The RNA

obtained at this stage was basically inhibitor-free and able to be used in RT-PCR.

In preliminary experiments, Norwalk virus RNA was detected in less than 0.05g of stool samples using the CTAB technique. A trace inhibitor activity was observed with RNA extracted with either CTAB or oligo d(T) alone, but this was easily removed by dilution (1:2) of the extracted nucleic acid before RT-PCR. Combination of the CTAB and oligo d(T) techniques resulted in obtaining high quality, inhibitor free RNA which could be used directly for RT-PCR detection and for cloning of the viral genome. With development of this method to clone from small amounts of stool, one skilled in the art will know that they can obtain cDNAs for the remainder of the genome including those representing the 5'-end of the genome.

For detection with PCR, primers based on the above nucleotide sequence of the genome were made by chemical methods. These primers include: Primer 1: CACGCGGAGGCTCTCAAT located at nucleotides 7448 to 7465; Primer 4: GGTGGCGAAGCGGCCCTC located at nucleotides 7010 to 7027; Primer 8: TCAGCAGTTATAGATATG located at nucleotides 1409 to 1426; Primer 9: ATGCTATATACATAGGTC located at nucleotides 612 to 629; Primer 16: CAACAGGTACTACGTGAC located at nucleotides 4010 to 4027; and Primer 17: TGTGGCCCAAGATTTGCT located at nucleotides 4654 to 4671 (SEQ ID NOS 51 through 56, respectively). These primers have been shown to be useful to detect virus using reverse transcription and polymerase chain reaction methods (RT-PCR). Figure 7 shows data using these primers. In primer sets 1 and 4, 8 and 9, and 16 and 17, the reverse compliments for the sequences given above for primers 1, 8, and 17 were used.

New, additional primer sets (Table 6 and SEQ ID NOS.: 15 to 37) are used as probes to detect the Norwalk-related viruses. Table 7 shows the ability of newly selected primer sets 36-35, 69-39, 78-80 to detect many Norwalk-related viruses. These results are additional examples of the use of primer sets from the original Norwalk virus sequence to detect Norwalk-related viruses. Nucleotide sequence data of many of these

viruses indicates that there is a continuum of genetic relatedness within the RNA region described by primer sets 36-35 or 69-39 of these different viruses (from 87% to 0%), yet these different agents can be detected using primers from the Norwalk virus genome sequence. The sequence of 2516
5 nucleotides of another small round structured virus (SRSV/KY/89 SEQ ID NO:12) also was obtained by using a total of 8 additional sets of primers from the original Norwalk virus sequence (primers 56 and 23, 42 and 55, 58 and 59, 60 and 61, 72 and 63, 76 and 77, 64 and 75, and 74 and 3; Table 6).

10

Example 6

Preparation of polyclonal antibodies

and monoclonal antibodies to Norwalk virus proteins

Protein(s) encoded in the cDNA fragments or derivatives thereof, is produced in a prokaryotic or eukaryotic expression system and used to
15 immunize animals to produce polyclonal antibodies for diagnostic assays. Prokaryotic hosts may include Gram negative as well as Gram positive bacteria, such as E. coli, S. typhimurium, Serratia marcescens, and Bacillus subtilis. Eukaryotic hosts may include yeast, insect or mammalian cells. Immunized animals may include mammals such as
20 guinea pigs, mice, rabbits, cows, goats or horses or other non-mammalian or non-murine species such as chickens. Repeated immunization of these animals with the expressed protein mixed with an adjuvant such as Freund adjuvant to enhance stimulation of an immune response produces antibodies to the protein.

25 Alternatively, synthetic peptides of greater than 15 amino acids made to match the amino acid sequence deduced from the partial cDNA sequence (or from other sequences determined by sequencing additional cDNAs detected with the original or other clones) are linked to a carrier protein such as bovine serum albumin or lysozyme or cross-linked with
30 treatment with glutaraldehyde and used to immunize animals to produce polyclonal antibodies for diagnostic tests.

The serum of animals immunized with either the expressed protein or with synthetic peptides are tested by immunologic assays such as immune electron microscopy, Western blots (immunoblots) and blocking ELISAs to demonstrate that antibodies to Norwalk and related viruses have been made. Reactivities with the expressed protein or synthetic peptides show specificity of the polyclonal sera. Reactivities with other viruses in the Norwalk group (Snow Mountain Agent, Hawaii Agent, Taunton Agent, etc.) indicate production of a reagent which recognizes cross-reacting epitopes.

Balb/c mice injected with the immunogens as described above and shown to have produced polyclonal antibodies are boosted with immunogen and then sacrificed. Their spleens are removed for fusion of splenocytes with myeloma cells to produce hybridomas. Hybridomas resulting from this fusion are screened for their reactivity with the expressed protein, the peptide and virus particles to select cells producing monoclonal antibodies to Norwalk virus. Screening of such hybridomas with Norwalk-related viruses permits identification of hybridomas secreting monoclonal antibodies to these viruses as well.

Development of Diagnostic Assays

Analysis of the deduced amino acid sequence of the Norwalk virus genome has shown that the Norwalk virus has the genetic organization shown in Figure 8. Expression of regions of this genome in cell-free translation systems and in the baculovirus expression system have shown that the 5'-end of the genome encodes nonstructural proteins and the 3'-end of the genome encodes at least one structural protein. Based on this information, one can express the complete genome or subgenomic regions of the genome to produce diagnostic assays to detect viral antigens or immune responses to specific regions of the genome. This information can be used to detect the Norwalk virus, antigens or immune responses to Norwalk virus. This information also can be used to detect other similar currently uncharacterized viruses that cause gastroenteritis or possibly other diseases. Some of these viruses will be in the Caliciviridae or in the

picornavirus superfamily. All of these viruses will have matching or similar genomic regions in their DNA sequences.

The availability of cDNA clones from viruses related to Norwalk virus enables the production of new antibodies and antisera for diagnostic assays for these related viruses. For example, availability of cDNA clones from caliciviruses which cannot be cultivated permits the expression of protein products of those clones. The protein products are used to develop new antibodies and antisera. In addition, genetic engineering is used to combine the cDNAs from viruses related to Norwalk virus with the cDNAs from Norwalk virus to produce chimeric proteins, such that part of the protein produced is derived from Norwalk virus genome sequence and another part of the protein is derived from the genome sequence of a virus related to Norwalk virus. These chimeric proteins are then used to produce diagnostic reagents, vaccines and antivirals. Examples of the diagnostic assays are shown in the specific examples and figures below.

Example 7

Development of diagnostic assays to detect nucleic acids
of Norwalk virus or Norwalk-related viruses by detection
of specific regions of the viral genomes
based on an understanding of the Norwalk virus genome.

The genetic organization of the Norwalk virus genome allows the prediction of specific regions of the gene sequence as regions where oligonucleotide primers or probes can be developed to detect Norwalk virus sequences and common sequences of other related or similar viruses. Some of these common genome sequences are found in viruses in the Caliciviridae or in the picornavirus superfamily. The detection can be done by standard PCR, hybridization or other gene amplification methods.

Two primers, named 35 (CTT GTT GGT TTG AGG CCA TAT, complementary to nt 4944-4924 in the Norwalk virus genome, SEQ ID NO: 15) and 36 (ATA AAA GTT GGC ATG AAC A, nt 4475-4493 in the Norwalk virus genome, SEQ ID NO: 16), were chosen from the region likely to encode the Norwalk virus RNA polymerase. These primers then

were used to prepare a cDNA clone by reverse transcriptase-PCR from the nucleotide sequence of human calicivirus Sapporo strain (HuCV Sapporo), 1982 outbreak (Figure 9, SEQ ID NO:5). The resulting sequence was compared to that of Norwalk virus and of feline and rabbit caliciviruses available from Genbank. The first cDNA clone from Sapporo, named "c-29_4-gel", determined to contain calicivirus sequence was 488 nucleotides long, of which 40 nucleotides were contributed by primers 36 and 35, leaving 448 nucleotides unique to human calicivirus Sapporo. The sequence of clone c-29_4-gel between primers 36 and 35 also is shown in Figure 9, SEQ ID NO:8.

Evidence that the HuCV Sapporo cDNA clone was correct is shown by five facts. First, the sequence exhibits strong homology with Norwalk virus, feline calicivirus, and the rabbit calicivirus at the nucleotide and amino acid levels. (See Figure 10 and Tables 7 and 8). Second, the sequence contains a continuous protein encoding region on the positive strand. In Norwalk, feline, and rabbit caliciviruses continuous protein encoding regions also are found in the region of homology. Third, the sequence contains the amino acid motif YGDD, which is a marker for RNA virus proteins which have RNA-dependent-RNA-polymerase activity. In c-29_4-gel, the YGDD motif is at the predicted distance from the ends of the sequence. Fourth, the same cDNA product was obtained from six different stool specimens. Fifth, no significant homologies were found for other sequences in the Genbank.

The nucleotide sequence of c-29_4-gel was used to synthesize an internal primer. This internal primer was used to prepare a second set of RT-PCR products from human calicivirus Sapporo RNA. A number of new cDNA clones were obtained of which one, named "at23s2m31", contains overlapping sequence which is 5' on the virus genome from that contained in c-29_4-gel. Sequence at23s2m31 is 149 nucleotides long (SEQ ID NO:7) and overlaps c-29_4-gel by 46 nucleotides. See Figure 9 for at23s2m31 sequence and area of overlap with c-29_4-gel. The resulting combined sequence information of c-29_4-gel and at23s2m31 is

551 nucleotides in length, excluding the portion c-29_4-gel contributed by prime 35.

Although the human calicivirus Sapporo sequence was generated from knowledge of the Norwalk virus sequence, the former is distinguishable in the same region (see Table 8 or Figure 9). The known sequence of human calicivirus Sapporo indicates that this virus is more closely related to the animal caliciviruses than to Norwalk virus.

In May, 1987, a child in Houston was infected with a virus which was identified as a calicivirus based on its morphology. Samples containing virus particles from this child failed to react in serologic assays developed for the detection of Norwalk virus and human calicivirus Sapporo. Primers 36 and 35 were used to prepare cDNA from the viral genome of this strain using RT-PCR. The resulting cDNA product, called 4847 complete, is 434 nucleotides long, excluding the primers, and is distinguishable from that of Norwalk virus and human calicivirus Sapporo. (See "Houston" in homology comparison in Figure 10; Table 10 and SEQ ID NO:10). Evidence that this Houston cDNA is correct is the same as that listed for c-29_4-gel above, except that homology with Norwalk virus and human calicivirus Sapporo is not statistically significant.

Use of the sequence from the human calicivirus
Sapporo strain to produce an amplification primers
for human calicivirus Sapporo and related agents

The known sequence of human calicivirus Sapporo overlaps one of the two primers, called primer 36 (see Table 6), used for the initial amplification of cDNA clone c-29_4-gel. Examination of the homology of known calicivirus sequences (Table 8 SEQ ID NOS 57 through 62) in that region indicated that a new 36 primer could be synthesized and used to amplify caliciviruses more closely related to human calicivirus Sapporo than Norwalk virus. A new primer was synthesized and is called primer "new 36" (see Table 6, last line, and SEQ ID NO:37).

The new 36 primer was used with primer 35 to generate a cDNA clone from a calicivirus which caused a diarrhea outbreak in November, 1986, in a Houston day care center ("Day care"). The calicivirus strain causing this Day care outbreak was antigenically related to human calicivirus Sapporo but antigenically distinct from Norwalk virus by EIA. The Day care cDNA product obtained from the RT-PCR reaction with primers new 36 and 35 is 445 nucleotides long, excluding the primers (see Figure 9 and SEQ ID NO:9), and has close homology to human calicivirus Sapporo and a more distant, yet still significant homology with Norwalk virus, as shown in Figure 10. Evidence that this Day care cDNA is correct is the same as that listed for c-29_4-gel above.

Use of primers 35 and 36 derived from the Norwalk virus sequence to derive a cDNA clone from an animal calicivirus

A calicivirus was isolated from the mouth of the pygmy chimpanzee, *Pan paniscus*. This calicivirus is antigenically distinct from the human calicivirus Sapporo strain by EIA. A cDNA was produced from the primate calicivirus (PrCV) RNA using RT-PCR and primers 36 and 35. The complete nucleotide sequence of this cDNA is not yet available. The cDNA, called atprcvw2 (Figure 20; SEQ. ID. NOS 13 and 14), is of the predicted size and has significant nucleotide homology with human calicivirus Sapporo, feline calicivirus(es), and the rabbit calicivirus in the region of known sequence. No significant homology with Norwalk virus has been observed in the region of known sequence. The known amino acid sequence contains the YGDD motif on the positive strand at the predicted distance from primer 35.

Use of multiple primers from the Norwalk virus genomic sequence to detect and characterize KY89, another small round virus associated with an outbreak of gastroenteritis.

The known sequence for Norwalk virus is used to obtain the sequence of other viruses such as SRSV/KY/89, an agent from a stool from an outbreak of gastroenteritis in Japan in 1989. Originally, cDNA

products and sequence information were obtained using primer sets 36-35. Continued work with another 8 sets of primers (Primers 56 and 23, 42 and 55, 58 and 59, 60 and 61, 72 and 63, 76 and 77, 64 and 75, and 74 and 3 in Table 6 and SEQ ID NOS:21 through 36) allowed the SRSV/KY/89
5 sequence of 2516 nucleotides to be determined (Figures 11 and 12, SEQ ID NO:12). This sequence includes the part of the polymerase region and the capsid region of the genome. Figures 14 and 6 (SEQ ID NOS 38 through 50 and 63 through 75) show sequences from other Norwalk-related viruses. Continued use of this approach with other Norwalk-
10 related viruses (such as those shown in Table 7) allows the discovery of the complete sequences of multiple Norwalk-related viruses. Those skilled in the art will realize that the use of such sequence information and expression of fragments and derivatives of Norwalk-related viruses permits development of diagnostic assays to detect antibodies, antigens,
15 viral genetic material or antivirals and to develop vaccines for specific Norwalk-related viruses in the same manner that Norwalk virus fragments and derivatives have been used.

Example 8

20 Development of diagnostic assays using expressed Norwalk virus proteins to detect immune responses to Norwalk virus

Protein(s) encoded in the Norwalk virus genome or fragments or derivatives thereof is produced in a prokaryotic or eukaryotic expression system and used as antigens in diagnostic assays to detect immune responses following virus infections. Prokaryotic hosts may include Gram
25 negative as well as Gram positive bacteria, such as *Escherichia coli*, *Salmonella typhimurium*, *Serratia marcescens*, *Bacillus subtilis*, *Staphylococcus aureus* and *Streptococcus sanguinis*. Eukaryotic hosts may include yeast, insect or mammalian cells. Diagnostic assays may include many formats such as enzyme-linked immunosorbent assays,
30 radioimmunoassays, immunoblots or other assays. Figure 15 shows data for a capsid protein encoded from the 3'-end of the Norwalk virus genome. It is expressed by nucleotides 5337 through 7753 of the DNA sequence

shown in Table 2 and Figure 8. This protein has an approximate molecular weight of 58,500 and is hereinafter referred to as the 58,500 mwt protein. It was produced in insect cells infected with baculovirus recombinants (C-6 and C-8). A band (see arrow in Figure 15) representing
5 the 58,500 mwt protein in C-6 and C-8 infected cells is not seen in insect cells infected with wild-type (WT) baculovirus or in mock infected cells. Other proteins encoded by Norwalk virus cDNA or fragments or derivatives are similarly expressed using baculovirus recombinants and other expression systems.

10 Figure 16 shows data using the 58,500 mwt protein produced using the baculovirus expression system to detect immune responses before and after infection of volunteers with Norwalk virus inoculum. Antigen was put on ELISA plates and pre- and post-infection human sera were added. The data show that when an individual has had the infection, the post-
15 serum reacts strongly to the antigen. Other proteins encoded in the Norwalk virus cDNA or fragments or derivatives thereof are similarly used to detect immune responses following Norwalk virus infection.

Some proteins have the intrinsic property of being able to form particles. The 58,500 mwt protein discussed above has that property.
20 Particles formed from proteins are expressed in any expression system and used to produce diagnostic assays based on detection of antibody responses or immune responses. Figure 17 shows an electron micrograph of particles produced using the baculovirus expression system from recombinants containing the 3'-end of the Norwalk genome. These
25 particles are similar in size to the native virus particles. They are antigenic, immunoreactive and immunogenic. They differ from most of the virus particles resulting from natural infection in that many of the expressed particles lack nucleic acids. The rNV particles are highly immunogenic when given parenterally to mice, rabbits and guinea pigs
30 and when given orally to mice.

Figure 18 shows data on the properties of rNV particles following centrifugation in gradients of CsCl. The density of the particles (symbolized by closed boxes) is 1.31 g/cc which is distinct from the 1.38

g/cc density of particles purified from the original infectious Norwalk inoculum given to volunteers. The gradients were fractionated. Each fraction was put on an ELISA plate and human serum was then introduced. The open boxes show that there was no ELISA activity with the pre-infection serum. The closed diamonds show there was reactivity with the post-infection serum. Other particles made from other proteins encoded in the Norwalk virus cDNA or fragments or derivatives thereof are similarly used to detect immune responses following Norwalk virus infection.

Figure 19 shows data using purified particles formed by the 58,500 mwt protein to detect immune responses in post-inoculation (but not pre-inoculation) serum samples of 9 volunteers infected with Norwalk virus. One of the volunteers, number 6, exhibited no symptoms of Norwalk virus infection based on monitoring clinical symptoms or measuring an immune response. Purified, expressed particles were put on ELISA plates and one pre- and one post-infection serum samples from each volunteer was added to the particles. The amount of antibody binding to the particles in each pre- and post-infection sample was measured. The data in Figure 19 show that the expressed proteins form particles that are immunoreactive and antigenic. Other proteins encoded in the Norwalk virus cDNA or fragments or derivatives thereof are similarly used to detect immunoreactive and antigenic activity.

Additional developments of diagnostic assays for the detection of Norwalk and Norwalk-related viruses also were pursued. First, new ELISA assays were made based on utilizing the Norwalk virus capsid protein that was engineered to be synthesized from a cDNA fragment that was deduced from the Norwalk virus cDNA sequence and then produced using the baculovirus expression system. This expressed Norwalk virus capsid protein self-assembled into recombinant Norwalk virus particles (rNV). Two new ELISA assays were established using this rNV antigen. One assay detects antiviral antibody and the other detects viral antigen. Both the ELISAs are very sensitive when compared to the previous assays (based on reagents from human volunteers) available to detect such

agents. Further characterization of the antibody ELISA has shown this assay detects immune responses following human infections with Norwalk virus and a subset of human infections with viruses in the Norwalk group such as Snow Mountain and Hawaii agents. In contrast, the antigen

5 ELISA is based on use of hyperimmune serum made to the baculovirus expressed recombinant Norwalk virus particles (rNV). This antigen ELISA has been found to be very specific in that it recognizes the prototype Norwalk virus (8FIIa) and a subset of closely related agents, but not all other viruses in the Norwalk group such as the Snow Mountain

10 agent and Hawaii agent (See Tables 1 and 7). While the antigen ELISA does not detect other viruses in the Norwalk group such as the small round structured viruses or caliciviruses, these and other Norwalk-related viruses have been able to be detected using primers selected from the nucleotide sequence of Norwalk virus (See Table 7).

15 To develop more broadly reactive diagnostic assays, ELISAs based on using other fragments of the Norwalk virus genome were developed. The new diagnostic assays are based on detection of antibody responses or of antigens deduced from fragments of the Norwalk virus genome other than the capsid region. An example and data of this approach is the

20 following.

One Norwalk virus nonstructural protein is predicted to be encoded in the first ORF of Norwalk viral genome. This ORF is located at the 5 end of the viral genome and it has a predicated molecular weight of 190,000 (190K). Whether this ORF 1 is useful in diagnostic assays first

25 was evaluated by expressing the protein encoded in the full length viral RNA, and then synthesizing and testing the immunoreactivity of the encoded protein using a cell-free system. This was accomplished by in vitro transcription of a full length cDNA (pGNV-F) of the Norwalk viral genome cDNAs. This full-length cDNA was constructed by ligation of

30 subgenomic derivatives of the original Norwalk virus cDNAs shown in the physical map in Figure 5. The in vitro synthesized NV mRNAs next were examined for their ability to direct the synthesis of a Norwalk virus specific protein by cell-free translation in rabbit reticulocyte lysates in the

presence of ^{35}S methionine to produce a radiolabeled protein. The expressed proteins were analyzed by polyacrylamide gel electrophoresis (PAGE). A clear band of approximate molecular weight of 130,000 was observed in the sample containing the viral RNA but not in the negative control (without viral RNA). The immunoreactivity of this protein was examined by reactivity with pre- and post-infection sera from volunteers given Norwalk virus. The 130K protein was precipitated by a convalescent serum of a volunteer infected with Norwalk virus, but not by serum collected before infection, indicating this protein was virus-specific. This showed this 130K protein contains some immunoreactive epitopes. The apparent smaller size of the protein made in this translation system suggested that either the protein migrates aberrantly on gels, or an internal initiation codon was used to begin translation or some type of post translational modification may have occurred after the protein was translated.

To further characterize immunoreactive derivatives of the Norwalk virus cDNA useful for diagnostic assays, the 2C region of the Norwalk viral genome (see Figure 8) was expressed using the baculovirus expression system. This region was selected for initial expression because it is located at the 5'-end of the non-structural protein and a high level of conservation was found between the sequence of the predicted Norwalk virus protein, and new sequence published for related caliciviruses and picornavirus. A 5'-end cDNA fragment of the viral genome was subcloned into the baculovirus transfer vector pVL 1393. After co-transfection of insect Sf9 cells with wild-type baculovirus DNA, recombinants containing the Norwalk viral gene were identified and selected. After three rounds of plaque purification, radiolabeled lysates of recombinant-infected insect cells were prepared, and the radiolabeled proteins were analyzed by PAGE. The results showed that a protein of apparent molecular weight of 57,000 (57K) was made in recombinant-infected but not in uninfected cells. The size of the protein suggested that the internal AUG initiation codon located at nucleotide 953 was used for making this protein. This 57K protein also was precipitated by convalescent serum (but not by pre-

infection serum) from a volunteer who was infected with Norwalk virus. This protein mainly remained cell-associated. One skilled in the art will readily see that improvements in the yield and purification of this 2C nonstructural protein are possible and will yield more rapid ELISAs to
5 detect Norwalk and related virus infections. One skilled in the art also will see that by expressing proteins from other regions of the Norwalk viral genome (*e.g.*, 3C-like, 3D-like and the 3d ORF), diagnostic assays are made for Norwalk and related viruses similar to the ELISAs made with the 2C nonstructural and rNV structural protein. These new assays
10 should widen the spectrum in detection of Norwalk-related viruses.

The initial lack of sensitive methods to detect Norwalk and Norwalk-related viruses made the description of the many Norwalk-related viruses difficult to define. However, as shown in Table 7, the methods and data provided here demonstrate how the discovery of the nucleotide sequence
15 of the Norwalk virus genome has led to the ability to develop tests to detect Norwalk virus and other related agents. The data and methods also demonstrate that fragments and derivatives of the Norwalk virus genome can be used to provide evidence of and immunity against Norwalk and related viruses.

20

Example 9

Development of diagnostic assays using expressed Norwalk virus and Norwalk-related viruses to detect viral antigens

Individual proteins, particles or protein aggregates formed from expression of one or more Norwalk virus genes in any prokaryotic or
25 eukaryotic expression system are used as an immunogen or inoculate animals to produce polyclonal and monoclonal antibodies for diagnostic assays to detect viral antigens.

Recombinant Norwalk virus particles (rNV) produced using the baculovirus expression system has been used to produce polyclonal
30 antibodies in mice, guinea pigs and rabbits following parenteral immunization (see Table 9). Mice given rNV orally also have developed serum antibodies. Hybridomas from mice immunized with rNV also have

been obtain following fusion with myeloma cells. Use of these antibodies in a capture ELISA has shown NV antigen can be detected. This antigen ELISA based on the antiserum made to the rNV particles is quite specific and it detects only a subset of Norwalk-related viruses (See Table 7).

- 5 Therefore, additional capsid antigens from other Norwalk-related viruses (such a Snow Mountain, Hawaii etc.) must be expressed to produce a more broadly reactive ELISA for capsid antigen. The ELISA is only one format that can be used to detect virus antigen. Other formats could include immunofluorescence or immunocytochemistry, or immune electron
10 microscopy. The comparison of the capsid sequences of Norwalk virus and Norwalk-related viruses permits the identification of conserved regions of the capsid protein and use of fragments of such sequences to immunize animals and can result in the production of antisera with more broad reactivity to Norwalk-related viruses. Alternatively, sequential
15 immunization of animals with expressed proteins of Norwalk and Norwalk-related viruses will result in antiserum with the desired broad reactivity. Antigen detection assays that are specific to one of a few strains of Norwalk and Norwalk-related viruses and additional assays that are more broadly reactive each will have use.

- 20 Expression of fragments of proteins encoded in other regions of the genome can be used to produce antiserum to other proteins for use in ELISAs to detect viral antigens. The expression of the first ORF that represents a polyprotein encoded in the 5'-end of the genome and fragment 2C of the polyprotein has shown that each of these
25 nonstructural proteins in immunoreactive and antiserum made to these can be used to develop diagnostic assays to detect these viral proteins. These assays can be broadly reactive and detect many other Norwalk-related viruses because of sequence conservation. Those skilled in the art will recognize that knowledge of the genome organization of Norwalk
30 virus permits similar expression of the same regions of the genomes of other Norwalk-related viruses for use in diagnostic assays to detect viral antigens.

Example 10

Development of a vaccine using
Norwalk virus expressed antigens

Vaccines for Norwalk virus, the Norwalk group of viruses or other
5 small round viruses are made from an expressed Norwalk virus protein.
That expressed protein can be a Norwalk virus capsid protein expressed
alone or in combination with one or more other Norwalk virus proteins
or self-forming particles. For example, the particles shown in Figure 17
were produced using the baculovirus expression system. They are used as
10 a vaccine when expressed alone or in combination with one or more other
Norwalk virus proteins. Similarly, the other proteins encoded in the
Norwalk virus cDNA or fragments or derivatives thereof are used as a
vaccine when expressed alone or in combination with one or more
Norwalk virus proteins.

15 Individuals are vaccinated orally, parenterally or by a combination
of both methods. For parenteral vaccination, the expressed protein is
mixed with an adjuvant and administered in one or more doses in
amounts and at intervals that give maximum immune response and
protective immunity. Oral vaccination parallels natural infection by
20 Norwalk virus inoculum, i.e. the individual ingests the vaccine with
dechlorinated water or buffer. Oral vaccination may follow sodium
bicarbonate treatment to neutralize stomach activity. For example,
sodium bicarbonate solution is taken by each person 2 minutes before and
5 minutes after vaccine administration.

25

Example 11

Production of a vaccine for other agents by using
expressed Norwalk virus capsids as a carrier or vehicle
for the expression of other antigens
or parts of other antigens

30 Identification of the region of the genome that encodes the Norwalk
virus capsid protein and that forms particles following expression (i.e.,
regions 5346 through 6935 and 5337 through 7753) allows genetic

engineering of the cDNA that encodes the capsid protein to incorporate one or more heterologous pieces of cDNA that encode antigenic epitopes. Expression of such recombinant genes produces a recombinant capsid that is antigenic, induces antibodies, and protects against Norwalk virus and its antigens, and against the heterologous epitopes or antigens.

Alternatively, the Norwalk virus capsid protein carrier is mixed with or covalently linked to one or more heterologous protein antigens or synthetic peptides containing heterologous epitopes. This mixture is antigenic, induces antibodies, and protects against Norwalk virus and its antigens, and against the heterologous epitopes or antigens.

Individuals are vaccinated using the oral and parenteral methods described above in example 10.

Example 12

Kit

Kits for detecting immune responses to Norwalk virus and are prepared by supplying in a container a protein deduced from the Norwalk virus genome shown in Table 2 or fragments or derivatives thereof. Similar proteins are prepared from Norwalk-related viruses to detect immune responses to the Norwalk-related viruses. For example, the protein encoded by Norwalk virus nucleotides 1 through 7753, the protein encoded by Norwalk virus nucleotides 146 through 5359, the protein encoded by Norwalk virus nucleotides 5337 through 7573, the protein encoded by Norwalk virus nucleotides 5346 through 6935, the protein encoded by Norwalk virus nucleotides 6938 through 7573 and any combinations thereof may be used in such kits. The kit can also include controls for false positive and false negatives, reagents and sample collection devices. The kit can be equipped to detect one sample or multiple samples.

Example 13

Kit

Kits for detecting Norwalk viruses and Norwalk-related viruses are prepared by supplying in a container at least one antiserum made from a protein expressed from the deduced amino acid sequence of the Norwalk virus genome shown in Tables 3, 4, or 5 or from a fragment or derivative the deduced amino acid sequence. Similar antiserum are made from proteins encoded by Norwalk-related viruse genomes. For example, an antiserum made to the protein encoded by Norwalk virus nucleotides 1 through 7753, the protein encoded by Norwalk virus nucleotides 146 through 5359, the protein encoded by Norwalk virus nucleotides 5337 through 7573, the protein encoded by Norwalk virus nucleotides 5346 through 6935, the protein encoded by Norwalk virus nucleotides 6938 through 7573 and any combination thereof may be used in such kits. The kit can also include controls for false positives and false negatives, reagents and sample collection devices. The kit can be equipped to detect one sample or multiple samples.

In conclusion, it is seen that the present invention and the embodiments disclosed herein are well adapted to carry out the objectives and attain the ends and advantages mentioned as well as other inherent therein. The novel features characteristic of this invention are set forth in the appended claims. While presently preferred embodiments of the invention have been described for the purpose of disclosure, numerous changes in the details of synthesis and use described herein will be apparent to those skilled in the art. It should be understood, however, that there is no intention to limit the invention to the specific form disclosed, but on the contrary, the intention is to cover all modifications, alternative means of synthesis and use and equivalents falling within the spirit and scope of the invention.

Table 1. Classification of small round viruses.

Featureless viruses*

<u>Virus</u> Enterovirus	<u>Physical features</u> RNA BD ² 1.34 g/cm ³ Size range 20-30 nm	<u>Examples</u> Polio Hepatitis A
Parvovirus	DNA BD 1.38-1.46 g/cm ³ Size range 18-26 nm	Feline/mink/canine Bovine
Candidate parvovirus	DNA? BD 1.38-1.4g/cm ³	Wollan, Ditchling, Parramatta, cockle

Structured viruses**

<u>Virus</u> Astrovirus	<u>Morphology</u> 5--6-pointed surface star	<u>Physical features</u> RNA BD 1.36-1.38 g/cm ³ Size range 28-30 nm	<u>Examples</u> Lamb Human
Calicivirus	Surface hollows, ragged outline 'Star of David' configuration	RNA BD 1.36-1.39 g/cm ³ Size range 30-38 nm	Human (Norwalk, UK1-4, and Sapporo and other Japanese strains) Newbury (bovine) Pig
Small round structured virus (SRSV)	Amorphous surface, ragged outline	BD 1.36-1.41 g/cm ³ Size range 30-35 nm	Montgomery County, Hawaii, Taunton, Amulree, Otofuke, Snow Mountain

* Smooth outer edge and no discernible surface structure.

** Surface structure and/or ragged outline

2 BD = buoyant density

Table 2. The nucleotide sequence of Norwalk virus genome.

GGCGTCAAAA	GACGTCGTTC	CTACTGCTGC	TAGCAGTGAA	AATGCTAACA	ACAATAGTAG	60
TATTAAGTCT	CGTCTATTGG	CGAGACTCAA	GGGTTCAAGT	GGGGCTACGT	CCCCACCCAA	120
CTCGATAAAG	ATAACCAACC	AAGATATGGC	TCTGGGGCTG	ATTGGACAGG	TCCCAGCGCC	180
AAAGGCCACA	TCCGTCGATG	TCCCTAAACA	ACAGAGGGAT	AGACCACCAC	GGACTGTTGC	240
CGAAGTTCAA	CAAAATTTCG	GTTGGACTGA	GAGACCACAA	GACCAGAATG	TTAAGACGTG	300
GGATGAGCTT	GACCACACAA	CAAAACAACA	GATACTTGAT	GAACACGCTG	AGTGGTTTGA	360
TGCCGGTGGC	TTAGGTCCAA	GTACACTACC	CACTAGTCAT	GAACGGTACA	CACATGAGAA	420
TGATGAAGGC	CACCAGGTAA	AGTGGTCGGC	TAGGGAAGGT	GTAGACCTTG	GCATATCCGG	480
GCTCACGACG	GTGTCTGGGC	CTGAGTGGA	TATGTGCCCG	CTACCACCAG	TTGACCAAAG	540
GAGCAGCACA	CCTGCAACTG	AGCCCACAAT	TGGTGACATG	ATCGAATTCT	ATGAAGGGCA	600
CATCTATCAT	TATGCTATAT	ACATAGGTCA	AGGCAAGACG	GTGGGTGTAC	ACTCCCCTCA	660
AGCAGCCTTC	TCAATAACGA	GGATCACCAT	ACAGCCCAT	TCAGCTTGGT	GGCGAGTCTG	720
TTATGTCCCA	CAACCAAAAC	AGAGGCTCAC	ATACGACCAA	CTCAAAGAAT	TAGAAAATGA	780
ACCATGGCCG	TATGCCGCAG	TCACGAACAA	CTGCTTCGAA	TTTTGTTGCC	AGGTCATGTG	840
CTTGAAGAT	ACTTGGTTGC	AAAGGAAGCT	CATCTCCTCT	GGCCGGTTTT	ACCACCCGAC	900
CCAAGATTGG	TCCCAGAGACA	CTCCAGAATT	CCAACAAGAC	AGCAAGTTAG	AGATGGTTAG	960
GGATGCAGTG	CTAGCCGCTA	TAAATGGGTT	GGTGTGCGGG	CCATTTAAAG	ATCTTCTGGG	1020
TAAGCTCAAA	CCCTTGAACG	TGCTTAACTT	ACTTTCAAAC	TGTGATTGGA	CGTTCATGGG	1080
GGTCGTGGAG	ATGGTGGTCC	TCCTTTTAGA	ACTCTTTGGA	ATCTTTTGGG	ACCCACCTGA	1140
TGTTTCCAAC	TTTATAGCTT	CACTCCTGCC	AGATTTCCAT	CTACAGGGCC	CCGAGGACCT	1200
TGCCAGGGAT	CTCGTGCCAA	TAGTATTGGG	GGGGATCGGC	TTAGCCATAG	GATTCACCAG	1260
AGACAAGGTA	AGTAAGATGA	TGAAGAATGC	TGTTGATGGA	CTTGTGCGG	CAACCCAGCT	1320
CGGTCAATAT	GGCCTAGAAA	TATTCTCATT	ACTAAAGAAG	TACTTCTTCG	GTGGTGATCA	1380
AACAGAGAAA	ACCCTAAAAG	ATATTGAGTC	AGCAGTTATA	GATATGGAAG	TACTATCATC	1440
TACATCAGTG	ACTCAGCTCG	TGAGGGACAA	ACAGTCTGCA	CGGGCTTATA	TGGCCATCTT	1500
AGATAATGAA	GAAGAAAAGG	CAAGGAAATT	ATCTGTCAAG	AATGCCGACC	CACACGTAGT	1560
ATCCTCTACC	AATGCTCTCA	TATCCCGGAT	CTCAATGGCT	AGGGCTGCAT	TGGCCAAGGC	1620
TCAAGCTGAA	ATGACCAGCA	GGATGCGTCC	TGTGGTCATT	ATGATGTGTG	GGCCCCCTGG	1680
TATAGGTAAA	ACCAAGGCAG	CAGAACATCT	GGCTAAACGC	CTAGCCAATG	AGATACGGCC	1740
TGGTGGTAAG	GTTGGGCTGG	TCCCACGGGA	GGCAGTGGAT	CATTGGGATG	GATATCACGG	1800
AGAGGAAGTG	ATGCTGTGGG	ACGACTATGG	AATGACAAAG	ATACAGGAAG	ACTGTAATAA	1860
ACTGCAAGCC	ATAGCCGACT	CAGCCCCCCT	AACACTCAAT	TGTGACCGAA	TAGAAAACAA	1920
GGGAATGCAA	TTTGTGTCTG	ATGCTATAGT	CATCACCACC	AATGCTCCTG	GCCCAGCCCC	1980
AGTGGACTTT	GTCAACCTCG	GGCCTGTTTG	CCGAAGGGTG	GACTTCCTTG	TGTATTGCAC	2040
GGCACCTGAA	GTTGAACACA	CGAGGAAAGT	CAGTCCTGGG	GACACAACCTG	CACTGAAAGA	2100
CTGCTTCAAG	CCCGATTTC	CACATCTAAA	AATGGAGTTG	GCTCCCCAAG	GGGGCTTTGA	2160
TAACCAAGGG	AATACCCCGT	TTGGTAAGGG	TGTGATGAAG	CCCACCACCA	TAAACAGGCT	2220
GTTAATCCAG	GCTGTAGCCT	TGACGATGGA	GAGACAGGAT	GAGTTCCAAC	TCCAGGGGCC	2280
TACGTATGAC	TTTGATACTG	ACAGAGTAGC	TGCGTTTACG	AGGATGGCCC	GAGCCAACGG	2340
GTTGGGTCTC	ATATCCATGG	CCTCCCTAGG	CAAAAAGCTA	CGCAGTGTCA	CCACTATTGA	2400
AGGATTAAAG	AATGCTCTAT	CAGGCTATAA	AATATCAAAA	TGCAGTATAC	AATGGCAGTC	2460
AAGGGTGTAC	ATTATAGAAT	CAGATGGTGC	CAGTGTACAA	ATCAAAGAAG	ACAAGCAAGC	2520

Table 2, continued

TTTGACCCCT	CTGCAGCAGA	CAATTAACAC	GGCCTCACTT	GCCATCACTC	GACTCAAAGC	2580
AGCTAGGGCT	GTGGCATACG	CTTCATGTTT	CCAGTCCGCC	ATAACTACCA	TACTACAAAT	2640
GGCGGGATCT	GCGCTCGTTA	TTAATCGAGC	GGTCAAGCGT	ATGTTTGGTA	CCCGTACAGC	2700
AGCCATGGCA	TTAGAAGGAC	CTGGGAAAGA	ACATAATTGC	AGGGTCCATA	AGGCTAAGGA	2760
AGCTGGAAG	GGGCCCATAG	GTCATGATGA	CATGGTAGAA	AGGTTTGGCC	TATGTGAAAC	2820
TGAAGAGGAG	GAGAGTGAGG	ACCAAATTCA	AATGGTACCA	AGTGATGCCG	TCCCAGAAGG	2880
AAAGAACAAA	GGCAAGACCA	AAAAGGGACG	TGGTCGCAAA	AATAACTATA	ATGCATTCTC	2940
TCGCCGTGGT	CTGAGTGATG	AAGAATATGA	AGAGTACAAA	AAGATCAGAG	AAGAAAAGAA	3000
TGGCAATTAT	AGTATACAAG	AATACTTGGA	GGACCGCCAA	CGATATGAGG	AAGAATTAGC	3060
AGAGGTACAG	GCAGGTGGTG	ATGGTGGCAT	AGGAGAAACT	GAAATGAAA	TCCGTACACAG	3120
GGTCTTCTAT	AAATCCAAGA	GTAAGAAACA	CCAACAAGAG	CAACGGCGAC	AACTTGGTCT	3180
AGTGACTGGA	TCAGACATCA	GAAAACGTAA	GCCCATTGAC	TGGACCCCGC	CAAAGAATGA	3240
ATGGGCAGAT	GATGACAGAG	AGGTGGATTA	TAATGAAAAG	ATCAATTTTG	AAGCTCCCCC	3300
GACACTATGG	AGCCGAGTCA	CAAAGTTTGG	ATCAGGATGG	GGCTTTTGGG	TCAGCCCGAC	3360
AGTGTTTCATC	ACAACCACAC	ATGTAGTGCC	AACTGGTGTG	AAAGAATTCT	TTGGTGAGCC	3420
CCTATCTAGT	ATAGCAATCC	ACCAAGCAGG	TGAGTTCACA	CAATTCAGGT	TCTCAAAGAA	3480
AATGCGCCCT	GACTTGACAG	GTATGGTCCT	TGAAGAAGGT	TGCCCTGAAG	GGACAGTCTG	3540
CTCAGTCCTA	ATTAAACGGG	ATTCGGGTGA	ACTACTTCCG	CTAGCCGTCC	GTATGGGGGC	3600
TATTGCCCTCC	ATGAGGATAC	AGGGTCGGCT	TGTCCATGGC	CAATCAGGGA	TGTTACTGAC	3660
AGGGGCCAAT	GCAAAGGGGA	TGGATCTTGG	CACTATACCA	GGAGACTGCG	GGGCACCATA	3720
CGTCCACAAG	CGCGGAATG	ACTGGGTGT	GTGTGGAGTC	CACGCTGCAG	CCACAAAGTC	3780
AGGCAACACC	GTGGTCTGCG	CTGTACAGGC	TGGAGAGGGC	GAAACCGCAC	TAGAAGGTGG	3840
AGACAAGGGG	CATTATGCCG	GCCACGAGAT	TGTGAGGTAT	GGAAGTGGCC	CAGCACTGTC	3900
AACTAAAACA	AAATTCTGGA	GGTCCTCCCC	AGAACCACTG	CCCCCGGAG	TATATGAGCC	3960
AGCATACCTG	GGGGGCAAGG	ACCCCGTGT	ACAGAATGGC	CCATCCCTAC	AACAGGTACT	4020
ACGTGACCAA	CTGAAACCTT	TGCGGACCC	CGCGGCGCG	ATGCCTGAGC	CTGGCCTACT	4080
GGAGGCTGCG	GTTGAGACTG	TAACATCCAT	GTTAGAACAG	ACAATGGATA	CCCCAAGCCC	4140
GTGGTCTTAC	GCTGATGCCT	GCCAATCTCT	TGACAAAAC	ACTAGTTCGG	GGTACCCTCA	4200
CCATAAAAGG	AAGAATGATG	ATTGGAATGG	CACCACCTTC	GTTGGAGAGC	TCGGTGAGCA	4260
AGCTGCACAC	GCCAACAATA	TGTATGAGAA	TGCTAAACAT	ATGAAACCCA	TTTACACTGC	4320
AGCCTTAAAA	GATGAAGTAG	TCAAGCCAGA	AAAGATTAT	CAAAAAGTCA	AGAAGCGTCT	4380
ACTATGGGGC	GCCGATCTCG	GAACAGTGGT	CAGGGCCGCC	CGGGCTTTTG	GCCCATTTTG	4440
TGACGCTATA	AAATCACATG	TCATCAAATT	GCCAATAAAA	GTTGGCATGA	ACACAATAGA	4500
AGATGGCCCC	CTCATCTATG	CTGAGCATGC	TAAATATAAG	AATCATTTTG	ATGCAGATTA	4560
TACAGCATGG	GACTCAACAC	AAAATAGACA	AATTATGACA	GAATCCTTCT	CCATTATGTC	4620
GCGCCTTACG	GCCTCACCAG	AATTGGCCGA	GGTTGTGGCC	CAAGATTTCG	TAGCACCATC	4680
TGAGATGGAT	GTAGGTGATT	ATGTCATCAG	GGTCAAAGAG	GGGCTGCCAT	CTGGATTCCC	4740
ATGTACTTCC	CAGGTGAACA	GCATAAATCA	CTGGATAATT	ACTCTCTGTG	CACTGTCTGA	4800
GGCCACTGGT	TTATCACCTG	ATGTGGTGCA	ATCCATGTCA	TATTTCTCAT	TTTATGGTGA	4860
TGATGAGATT	GTGTCAACTG	ACATAGATT	TGACCCAGCC	CGCCTCACTC	AAATTCTCAA	4920
GGAATATGGC	CTCAAACCAA	CAAGGCCTGA	CAAAACAGAA	GGACCAATAC	AAGTGAGGAA	4980
AAATGTGGAT	GGACTGGTCT	TCTTGCGCG	CACCATTTCC	CGTGATGCGG	CAGGGTTCCA	5040
AGGCAGGTTA	GATAGGGCTT	CGATTGAACG	CCAAATCTTC	TGGACCCGCG	GGCCCAATCA	5100

Table 2, continued

TTCAGATCCA	TCAGAGACTC	TAGTGCCACA	CACTCAAAGA	AAAATACAGT	TGATTTCAC	5160
TCTAGGGGAA	GCTTCACTCC	ATGGTGAGAA	ATTTTACAGA	AAGATTTC	GCAAGGTCAT	5220
ACATGAAATC	AAGACTGGTG	GATTGGAAAT	GTATGTCCCA	GGATGGCAGG	CCATGTTCCG	5280
CTGGATGCGC	TTCCATGACC	TCGGATTGTG	GACAGGAGAT	CGCGATCTTC	TGCCCCGAATT	5340
CGTAAATGAT	GATGGCGTCT	AAGGACGCTA	CATCAAGCGT	GGATGGCGCT	AGTGGCGCTG	5400
GTCAGTTGGT	ACCGGAGGTT	AATGCTTCTG	ACCTCTTGC	AATGGATCCT	GTAGCAGGTT	5460
CTTCGACAGC	AGTCGCGACT	GCTGGACAAG	TTAATCCTAT	TGATCCCTGG	ATAATTAATA	5520
ATTTTGTGCA	AGCCCCCAA	GGTGAATTTA	CTATTTCCCC	AAATAATACC	CCCGGTGATG	5580
TTTTGTTTGA	TTTGAGTTTG	GGTCCCCATC	TTAATCCTTT	CTTGCTCCAT	CTATCACAAA	5640
TGTATAATGG	TTGGGTTGGT	AACATGAGAG	TCAGGATTAT	GCTAGCTGGT	AATGCCTTTA	5700
CTGCGGGGAA	GATAATAGTT	TCCTGCATAC	CCCCTGGTTT	TGGTTCACAT	AATCTTACTA	5760
TAGCACAAGC	AACTCTCTTT	CCACATGTGA	TTGCTGATGT	TAGGACTCTA	GACCCCATTG	5820
AGGTGCCTTT	GGAAGATGTT	AGGAATGTTT	TCCTTCATAA	TAATGATAGA	AATCAACAAA	5880
CCATGCGCCT	TGTGTGCATG	CTGTACACCC	CCCTCCGAC	TGGTGGTGGT	ACTGGTGATT	5940
CTTTTGTAGT	TGCAGGGCGA	GTTATGACTT	GCCCCAGTCC	TGATTTTAA	TTCTTGTTTT	6000
TAGTCCCTCC	TACGGTGGAG	CAGAAAACCA	GGCCCTTCAC	ACTCCCAAAT	CTGGCATTGA	6060
GTTCTCTGTC	TAACTCACGT	GCCCCTCTCC	CAATCAGTAG	TATGGGCATT	TCCCCAGACA	6120
ATGTCCAGAG	TGTGCAGTTC	CAAAATGGTC	GGTGTACTCT	GGATGGCCGC	CTGGTTGGCA	6180
CCACCCCACT	TTCATTGTCA	CATGTTGCCA	AGATAAGAGG	GACCTCCAAT	GGCACTGTAA	6240
TCAACCTTAC	TGAATTGGAT	GGCACACCC	TTACCCCTTT	TGAGGGCCCT	GCCCCCATTG	6300
GGTTTCCAGA	CCTCGGTGGT	TGTGATTGGC	ATATCAATAT	GACACAGTTT	GGCCATTCTA	6360
GCCAGACCCA	GTATGATGTA	GACACCACCC	CTGACACTTT	TGTCCCCCAT	CTTGGTTCAA	6420
TTCAGGCAAA	TGGCATTGGC	AGTGGTAATT	ATGTTGGTGT	TCTTAGCTGG	ATTTCCCCCC	6480
CATCACACCC	GTCTGGCTCC	CAAGTTGACC	TTTGGGAAGAT	CCCCAATTAT	GGGTCAAGTA	6540
TTACGGAGGC	AACACATCTA	GCCCCTTCTG	TATACCCCC	TGGTTTCGGA	GAGGTATTGG	6600
TCTTTTTCAT	GTCAAAAATG	CCAGGTCCTG	GTGCTTATAA	TTTGCCCTGT	CTATTACCAC	6660
AAGAGTACAT	TTCACATCTT	GCTAGTGAAC	AAGCCCTAC	TGTAGGTGAG	GCTGCCCTGC	6720
TCCACTATGT	TGACCCTGAT	ACCGGTCGGA	ATCTTGGGGA	ATTCAAAGCA	TACCCTGATG	6780
GTTTCCTCAC	TTGTGTCCCC	AATGGGGCTA	GCTCGGGTCC	ACAACAGCTG	CCGATCAATG	6840
GGGTCTTTGT	CTTTGTTTCA	TGGGTGTCCA	GATTTTATCA	ATTAAAGCCT	GTGGGAAGTG	6900
CCAGCTCGGC	AAGAGGTAGG	CTTGGTCTGC	GCCGATAATG	GCCCAAGCCA	TAATTGGTGC	6960
AATTGCTGCT	TCCACAGCAG	GTAGTGCTCT	GGGAGCGGGC	ATACAGGTG	GTGGCGAAGC	7020
GGCCCTCCAA	AGCCAAAGGT	ATCAACAAAA	TTTGCAACTG	CAAGAAAATT	CTTTTAAACA	7080
TGACAGGGAA	ATGATTGGGT	ATCAGGTTGA	AGCTTCAAAT	CAATTATTGG	CTAAAAATTT	7140
GGCAACTAGA	TATTCACCTC	TCCGTGCTGG	GGGTTTGACC	AGTGCTGATG	CAGCAAGATC	7200
TGTGGCAGGA	GCTCCAGTCA	CCCGCATTGT	AGATTGGAAT	GGCGTGAGAG	TGTCTGCTCC	7260
CGAGTCTCT	GCTACCACAT	TGAGATCCGG	TGGCTTCATG	TGAGTTCCCA	TACCATTGTC	7320
CTCTAAGCAA	AAACAGGTTT	AATCATCTGG	TATTAGTAAT	CCAAATTATT	CCCCTTCATC	7380
CATTCTCGA	ACCACTAGTT	GGGTCGAGTC	ACAAAACCTA	TCGAGATTG	GAAATCTTTC	7440
TCCATACCAC	GCGGAGGCTC	TCAATACAGT	GTGGTTGACT	CCACCCGGTT	CAACAGCCTC	7500
TTCTACACTG	TCTTCTGTGC	CACGTGGTTA	TTTCAATACA	GACAGGTTGC	CATTATTGCG	7560
AAATAATAGG	CGATGATGTT	GTAATATGAA	ATGTGGGCAT	CATATTCATT	TAATTAGGTT	7620
TAATTAGGTT	TAATTTGATG	TTAAAAAATA	AAAAAATAAA	AAAAAATAAA	AAAAAATAAA	7680

Tabl 2, continued

AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA AAAAAAAAAA	7740
AAAAAAAAAA AAA	7753

Table 3. The amino acid sequence deduced from nucleotides 146 through 5359 of the Norwalk virus genome shown in Table 2.

CTCGATAAAG ATAACCAACC AAGAT ATG GCT CTG GGG CTG ATT GGA CAG GTC	172
Met Ala Leu Gly Leu Ile Gly Gln Val	
1 5	
CCA GCG CCA AAG GCC ACA TCC GTC GAT GTC CCT AAA CAA CAG AGG GAT	220
Pro Ala Pro Lys Ala Thr Ser Val Asp Val Pro Lys Gln Gln Arg Asp	
10 15 20 25	
AGA CCA CCA CGG ACT GTT GCC GAA GTT CAA CAA AAT TTG CGT TGG ACT	268
Arg Pro Pro Arg Thr Val Ala Glu Val Gln Gln Asn Leu Arg Trp Thr	
30 35 40	
GAG AGA CCA CAA GAC CAG AAT GTT AAG ACG TGG GAT GAG CTT GAC CAC	316
Glu Arg Pro Gln Asp Gln Asn Val Lys Thr Trp Asp Glu Leu Asp His	
45 50 55	
ACA ACA AAA CAA CAG ATA CTT GAT GAA CAC GCT GAG TGG TTT GAT GCC	364
Thr Thr Lys Gln Gln Ile Leu Asp Glu His Ala Glu Trp Phe Asp Ala	
60 65 70	
GGT GGC TTA GGT CCA AGT ACA CTA CCC ACT AGT CAT GAA CGG TAC ACA	412
Gly Gly Leu Gly Pro Ser Thr Leu Pro Thr Ser His Glu Arg Tyr Thr	
75 80 85	
CAT GAG AAT GAT GAA GGC CAC CAG GTA AAG TGG TCG GCT AGG GAA GGT	460
His Glu Asn Asp Glu Gly His Gln Val Lys Trp Ser Ala Arg Glu Gly	
90 95 100 105	
GTA GAC CTT GGC ATA TCC GGG CTC ACG ACG GTG TCT GGG CCT GAG TGG	508
Val Asp Leu Gly Ile Ser Gly Leu Thr Val Ser Gly Pro Glu Trp	
110 115 120	
AAT ATG TGC CCG CTA CCA CCA GTT GAC CAA AGG AGC ACG ACA CCT GCA	556
Asn Met Cys Pro Leu Pro Pro Val Asp Gln Arg Ser Thr Thr Pro Ala	
125 130 135	
ACT GAG CCC ACA ATT GGT GAC ATG ATC GAA TTC TAT GAA GGG CAC ATC	604
Thr Glu Pro Thr Ile Gly Asp Met Ile Glu Phe Tyr Glu Gly His Ile	
140 145 150	
TAT CAT TAT GCT ATA TAC ATA GGT CAA GGC AAG ACG GTG GGT GTA CAC	652
Tyr His Tyr Ala Ile Tyr Ile Gly Gln Gly Lys Thr Val Gly Val His	
155 160 165	
TCC CCT CAA GCA GCC TTC TCA ATA ACG AGG ATC ACC ATA CAG CCC ATA	700
Ser Pro Gln Ala Ala Phe Ser Ile Thr Arg Ile Thr Ile Gln Pro Ile	
170 175 180 185	
TCA GCT TGG TGG CGA GTC TGT TAT GTC CCA CAA CCA AAA CAG AGG CTC	748
Ser Ala Trp Trp Arg Val Cys Tyr Val Pro Gln Pro Lys Gln Arg Leu	
190 195 200	
ACA TAC GAC CAA CTC AAA GAA TTA GAA AAT GAA CCA TGG CCG TAT GCC	796
Thr Tyr Asp Gln Leu Lys Glu Leu Glu Asn Glu Pro Trp Pro Tyr Ala	
205 210 215	
GCA GTC ACG AAC AAC TGC TTC GAA TTT TGT TGC CAG GTC ATG TGC TTG	844
Ala Val Thr Asn Asn Cys Phe Glu Phe Cys Cys Gln Val Met Cys Leu	
220 225 230	

Table 3, continued

GAA GAT ACT TGG TTG CAA AGG AAG CTC ATC TCC TCT GGC CGG TTT TAC	892
Glu Asp Thr Trp Leu Gln Arg Lys Leu Ile Ser Ser Gly Arg Phe Tyr	
235 240 245	
CAC CCG ACC CAA GAT TGG TCC CGA GAC ACT CCA GAA TTC CAA CAA GAC	940
His Pro Thr Gln Asp Trp Ser Arg Asp Thr Pro Glu Phe Gln Gln Asp	
250 255 260 265	
AGC AAG TTA GAG ATG GTT AGG GAT GCA GTG CTA GCC GCT ATA AAT GGG	988
Ser Lys Leu Glu Met Val Arg Asp Ala Val Leu Ala Ala Ile Asn Gly	
270 275 280	
TTG GTG TCG CGG CCA TTT AAA GAT CTT CTG GGT AAG CTC AAA CCC TTG	1036
Leu Val Ser Arg Pro Phe Lys Asp Leu Leu Gly Lys Leu Lys Pro Leu	
285 290 295	
AAC GTG CTT AAC TTA CTT TCA AAC TGT GAT TGG ACG TTC ATG GGG GTC	1084
Asn Val Leu Asn Leu Leu Ser Asn Cys Asp Trp Thr Phe Met Gly Val	
300 305 310	
GTG GAG ATG GTG GTC CTC CTT TTA GAA CTC TTT GGA ATC TTT TGG AAC	1132
Val Glu Met Val Val Leu Leu Glu Leu Phe Gly Ile Phe Trp Asn	
315 320 325	
CCA CCT GAT GTT TCC AAC TTT ATA GCT TCA CTC CTG CCA GAT TTC CAT	1180
Pro Pro Asp Val Ser Asn Phe Ile Ala Ser Leu Leu Pro Asp Phe His	
330 335 340 345	
CTA CAG GGC CCC GAG GAC CTT GCC AGG GAT CTC GTG CCA ATA GTA TTG	1228
Leu Gln Gly Pro Glu Asp Leu Ala Arg Asp Leu Val Pro Ile Val Leu	
350 355 360	
GGG GGG ATC GGC TTA GCC ATA GGA TTC ACC AGA GAC AAG GTA AGT AAG	1276
Gly Gly Ile Gly Leu Ala Ile Gly Phe Thr Arg Asp Lys Val Ser Lys	
365 370 375	
ATG ATG AAG AAT GCT GTT GAT GGA CTT CGT GCG GCA ACC CAG CTC GGT	1324
Met Met Lys Asn Ala Val Asp Gly Leu Arg Ala Ala Thr Gln Leu Gly	
380 385 390	
CAA TAT GGC CTA GAA ATA TTC TCA TTA CTA AAG AAG TAC TTC TTC GGT	1372
Gln Tyr Gly Leu Glu Ile Phe Ser Leu Leu Lys Lys Tyr Phe Phe Gly	
395 400 405	
GGT GAT CAA ACA GAG AAA ACC CTA AAA GAT ATT GAG TCA GCA GTT ATA	1420
Gly Asp Gln Thr Glu Lys Thr Leu Lys Asp Ile Glu Ser Ala Val Ile	
410 415 420 425	
GAT ATG GAA GTA CTA TCA TCT ACA TCA GTG ACT CAG CTC GTG AGG GAC	1468
Asp Met Glu Val Leu Ser Ser Thr Ser Val Thr Gln Leu Val Arg Asp	
430 435 440	
AAA CAG TCT GCA CGG GCT TAT ATG GCC ATC TTA GAT AAT GAA GAA GAA	1516
Lys Gln Ser Ala Arg Ala Tyr Met Ala Ile Leu Asp Asn Glu Glu Glu	
445 450 455	
AAG GCA AGG AAA TTA TCT GTC AGG AAT GCC GAC CCA CAC GTA GTA TCC	1564
Lys Ala Arg Lys Leu Ser Val Arg Asn Ala Asp Pro His Val Val Ser	
460 465 470	
TCT ACC AAT GCT CTC ATA TCC CGG ATC TCA ATG GCT AGG GCT GCA TTG	1612
Ser Thr Asn Ala Leu Ile Ser Arg Ile Ser Met Ala Arg Ala Ala Leu	
475 480 485	

Table 3, continued

GCC AAG GCT CAA GCT GAA ATG ACC AGC AGG ATG CGT CCT GTG GTC ATT	1660
Ala Lys Ala Gln Ala Glu Met Thr Ser Arg Met Arg Pro Val Val Ile	
490 495 500 505	
ATG ATG TGT GGG CCC CCT GGT ATA GGT AAA ACC AAG GCA GCA GAA CAT	1708
Met Met Cys Gly Pro Gly Ile Gly Lys Thr Lys Ala Ala Glu His	
510 515 520	
CTG GCT AAA CGC CTA GCC AAT GAG ATA CGG CCT GGT GGT AAG GTT GGG	1756
Leu Ala Lys Arg Leu Ala Asn Glu Ile Arg Pro Gly Gly Lys Val Gly	
525 530 535	
CTG GTC CCA CGG GAG GCA GTG GAT CAT TGG GAT GGA TAT CAC GGA GAG	1804
Leu Val Pro Arg Glu Ala Val Asp His Trp Asp Gly Tyr His Gly Glu	
540 545 550	
GAA GTG ATG CTG TGG GAC GAC TAT GGA ATG ACA AAG ATA CAG GAA GAC	1852
Glu Val Met Leu Trp Asp Tyr Gly Met Thr Lys Ile Gln Glu Asp	
555 560 565	
TGT AAT AAA CTG CAA GCC ATA GCC GAC TCA GCC CCC CTA ACA CTC AAT	1900
Cys Asn Lys Leu Gln Ala Ile Ala Asp Ser Ala Pro Leu Thr Leu Asn	
570 575 580 585	
TGT GAC CGA ATA GAA AAC AAG GGA ATG CAA TTT GTG TCT GAT GCT ATA	1948
Cys Asp Arg Ile Glu Asn Lys Gly Met Gln Phe Val Ser Asp Ala Ile	
590 595 600	
GTC ATC ACC ACC AAT GCT CCT GGC CCA GCC CCA GTG GAC TTT GTC AAC	1996
Val Ile Thr Asn Ala Pro Gly Pro Ala Pro Val Asp Phe Val Asn	
605 610 615	
CTC GGG CCT GTT TGC CGA AGG GTG GAC TTC CTT GTG TAT TGC ACG GCA	2044
Leu Gly Pro Val Cys Arg Arg Val Asp Phe Leu Val Tyr Cys Thr Ala	
620 625 630	
CCT GAA GTT GAA CAC ACG AGG AAA GTC AGT CCT GGG GAC ACA ACT GCA	2092
Pro Glu Val Glu His Thr Arg Lys Val Ser Pro Gly Asp Thr Thr Ala	
635 640 645	
CTG AAA GAC TGC TTC AAG CCC GAT TTC TCA CAT CTA AAA ATG GAG TTG	2140
Leu Lys Asp Cys Phe Lys Pro Asp Phe Ser His Leu Lys Met Glu Leu	
650 655 660 665	
GCT CCC CAA GGG GGC TTT GAT AAC CAA GGG AAT ACC CCG TTT GGT AAG	2188
Ala Pro Gln Gly Gly Phe Asp Asn Gln Gly Asn Thr Pro Phe Gly Lys	
670 675 680	
GGT GTG ATG AAG CCC ACC ACC ATA AAC AGG CTG TTA ATC CAG GCT GTA	2236
Gly Val Met Lys Pro Thr Thr Ile Asn Arg Leu Leu Ile Gln Ala Val	
685 690 695	
GCC TTG ACG ATG GAG AGA CAG GAT GAG TTC CAA CTC CAG GGG CCT ACG	2284
Ala Leu Thr Met Glu Arg Gln Asp Glu Phe Gln Leu Gln Gly Pro Thr	
700 705 710	
TAT GAC TTT GAT ACT GAC AGA GTA GCT GCG TTC ACG AGG ATG GCC CGA	2332
Tyr Asp Phe Asp Thr Asp Arg Val Ala Ala Phe Thr Arg Met Ala Arg	
715 720 725	
GCC AAC GGG TTG GGT CTC ATA TCC ATG GCC TCC CTA GGC AAA AAG CTA	2380
Ala Asn Gly Leu Gly Leu Ile Ser Met Ala Ser Leu Gly Lys Lys Leu	
730 735 740 745	

Table 3, continued

CGC	AGT	GTC	ACC	ACT	ATT	GAA	GGA	TTA	AAG	AAT	GCT	CTA	TCA	GGC	TAT	2428
Arg	Ser	Val	Thr	Thr	Ile	Glu	Gly	Leu	Lys	Asn	Ala	Leu	Ser	Gly	Tyr	
				750					755					760		
AAA	ATA	TCA	AAA	TGC	AGT	ATA	CAA	TGG	CAG	TCA	AGG	GTG	TAC	ATT	ATA	2476
Lys	Ile	Ser	Lys	Cys	Ser	Ile	Gln	Trp	Gln	Ser	Arg	Val	Tyr	Ile	Ile	
			765					770					775			
GAA	TCA	GAT	GGT	GCC	AGT	GTA	CAA	ATC	AAA	GAA	GAC	AAG	CAA	GCT	TTG	2524
Glu	Ser	Asp	Gly	Ala	Ser	Val	Gln	Ile	Lys	Glu	Asp	Lys	Gln	Ala	Leu	
		780					785					790				
ACC	CCT	CTG	CAG	CAG	ACA	ATT	AAC	ACG	GCC	TCA	CTT	GCC	ATC	ACT	CGA	2572
Thr	Pro	Leu	Gln	Gln	Thr	Ile	Asn	Thr	Ala	Ser	Leu	Ala	Ile	Thr	Arg	
	795					800					805					
CTC	AAA	GCA	GCT	AGG	GCT	GTG	GCA	TAC	GCT	TCA	TGT	TTC	CAG	TCC	GCC	2620
Leu	Lys	Ala	Ala	Arg	Ala	Val	Ala	Tyr	Ala	Ser	Cys	Phe	Gln	Ser	Ala	
810					815					820					825	
ATA	ACT	ACC	ATA	CTA	CAA	ATG	GCG	GGA	TCT	GCG	CTC	GTT	ATT	AAT	CGA	2668
Ile	Thr	Thr	Ile	Leu	Gln	Met	Ala	Gly	Ser	Ala	Leu	Val	Ile	Asn	Arg	
				830					835					840		
GCG	GTC	AAG	CGT	ATG	TTT	GGT	ACC	CGT	ACA	GCA	GCC	ATG	GCA	TTA	GAA	2716
Ala	Val	Lys	Arg	Met	Phe	Gly	Thr	Arg	Thr	Ala	Ala	Met	Ala	Leu	Glu	
			845					850					855			
GGA	CCT	GGG	AAA	GAA	CAT	AAT	TGC	AGG	GTC	CAT	AAG	GCT	AAG	GAA	GCT	2764
Gly	Pro	Gly	Lys	Glu	His	Asn	Cys	Arg	Val	His	Lys	Ala	Lys	Glu	Ala	
		860					865					870				
GGA	AAG	GGG	CCC	ATA	GGT	CAT	GAT	GAC	ATG	GTA	GAA	AGG	TTT	GGC	CTA	2812
Gly	Lys	Gly	Pro	Ile	Gly	His	Asp	Asp	Met	Val	Glu	Arg	Phe	Gly	Leu	
	875					880					885					
TGT	GAA	ACT	GAA	GAG	GAG	GAG	AGT	GAG	GAC	CAA	ATT	CAA	ATG	GTA	CCA	2860
Cys	Glu	Thr	Glu	Glu	Glu	Glu	Ser	Glu	Asp	Gln	Ile	Gln	Met	Val	Pro	
890					895					900					905	
AGT	GAT	GCC	GTC	CCA	GAA	GGA	AAG	AAC	AAA	GGC	AAG	ACC	AAA	AAG	GGA	2908
Ser	Asp	Ala	Val	Pro	Glu	Gly	Lys	Asn	Lys	Gly	Lys	Thr	Lys	Lys	Gly	
				910					915					920		
CGT	GGT	CGC	AAA	AAT	AAC	TAT	AAT	GCA	TTC	TCT	CGC	CGT	GGT	CTG	AGT	2956
Arg	Gly	Arg	Lys	Asn	Asn	Tyr	Asn	Ala	Phe	Ser	Arg	Arg	Gly	Leu	Ser	
			925					930					935			
GAT	GAA	GAA	TAT	GAA	GAG	TAC	AAA	AAG	ATC	AGA	GAA	GAA	AAG	AAT	GGC	3004
Asp	Glu	Glu	Tyr	Glu	Glu	Tyr	Lys	Lys	Ile	Arg	Glu	Glu	Lys	Asn	Gly	
		940					945					950				
AAT	TAT	AGT	ATA	CAA	GAA	TAC	TTG	GAG	GAC	CGC	CAA	CGA	TAT	GAG	GAA	3052
Asn	Tyr	Ser	Ile	Gln	Glu	Tyr	Leu	Glu	Asp	Arg	Gln	Arg	Tyr	Glu	Glu	
	955					960					965					
GAA	TTA	GCA	GAG	GTA	CAG	GCA	GGT	GGT	GAT	GGT	GGC	ATA	GGA	GAA	ACT	3100
Glu	Leu	Ala	Glu	Val	Gln	Ala	Gly	Gly	Asp	Gly	Gly	Ile	Gly	Glu	Thr	
970					975					980					985	
GAA	ATG	GAA	ATC	CGT	CAC	AGG	GTC	TTC	TAT	AAA	TCC	AAG	AGT	AAG	AAA	3148
Glu	Met	Glu	Ile	Arg	His	Arg	Val	Phe	Tyr	Lys	Ser	Lys	Ser	Lys	Lys	
				990					995						1000	

Table 3, continued

CAC CAA CAA GAG CAA CGG CGA CAA CTT GGT CTA GTG ACT GGA TCA GAC	3196
His Gln Gln Glu Gln Arg Arg Gln Leu Gly Leu Val Thr Gly Ser Asp	
1005 1010 1015	
ATC AGA AAA CGT AAG CCC ATT GAC TGG ACC CCG CCA AAG AAT GAA TGG	3244
Ile Arg Lys Arg Lys Pro Ile Asp Trp Thr Pro Pro Lys Asn Glu Trp	
1020 1025 1030	
GCA GAT GAT GAC AGA GAG GTG GAT TAT AAT GAA AAG ATC AAT TTT GAA	3292
Ala Asp Asp Asp Arg Glu Val Asp Tyr Asn Glu Lys Ile Asn Phe Glu	
1035 1040 1045	
GCT CCC CCG ACA CTA TGG AGC CGA GTC ACA AAG TTT GGA TCA GGA TGG	3340
Ala Pro Pro Thr Leu Trp Ser Arg Val Thr Lys Phe Gly Ser Gly Trp	
1050 1055 1060 1065	
GGC TTT TGG GTC AGC CCG ACA GTG TTC ATC ACA ACC ACA CAT GTA GTG	3388
Gly Phe Trp Val Lys Pro Thr Val Phe Ile Thr Thr Thr His Val Val	
1070 1075 1080	
CCA ACT GGT GTG AAA GAA TTC TTT GGT GAG CCC CTA TCT AGT ATA GCA	3436
Pro Thr Gly Val Lys Glu Phe Phe Gly Glu Pro Leu Ser Ser Ile Ala	
1085 1090 1095	
ATC CAC CAA GCA GGT GAG TTC ACA CAA TTC AGG TTC TCA AAG AAA ATG	3484
Ile His Gln Ala Gly Glu Phe Thr Gln Phe Arg Phe Ser Lys Lys Met	
1100 1105 1110	
CGC CCT GAC TTG ACA GGT ATG GTC CTT GAA GAA GGT TGC CCT GAA GGG	3532
Arg Pro Asp Leu Thr Gly Met Val Leu Glu Glu Gly Cys Pro Glu Gly	
1115 1120 1125	
ACA GTC TGC TCA GTC CTA ATT AAA CGG GAT TCG GGT GAA CTA CTT CCG	3580
Thr Val Cys Ser Val Leu Ile Lys Arg Asp Ser Gly Glu Leu Leu Pro	
1130 1135 1140 1145	
CTA GCC GTC CGT ATG GGG GCT ATT GCC TCC ATG AGG ATA CAG GGT CCG	3628
Leu Ala Val Arg Met Gly Ala Ile Ala Ser Met Arg Ile Gln Gly Arg	
1150 1155 1160	
CTT GTC CAT GGC CAA TCA GGG ATG TTA CTG ACA GGG GCC AAT GCA AAG	3676
Leu Val His Gly Gln Ser Gly Met Leu Leu Thr Gly Ala Asn Ala Lys	
1165 1170 1175	
GGG ATG GAT CTT GGC ACT ATA CCA GGA GAC TGC GGG GCA CCA TAC GTC	3724
Gly Met Asp Leu Gly Thr Ile Pro Gly Asp Cys Gly Ala Pro Tyr Val	
1180 1185 1190	
CAC AAG CGC GGG AAT GAC TGG GTT GTG TGT GGA GTC CAC GCT GCA GCC	3772
His Lys Arg Gly Asn Asp Trp Val Val Cys Gly Val His Ala Ala Ala	
1195 1200 1205	
ACA AAG TCA GGC AAC ACC GTG GTC TGC GCT GTA CAG GCT GGA GAG GGC	3820
Thr Lys Ser Gly Asn Thr Val Val Cys Ala Val Gln Ala Gly Glu Gly	
1210 1215 1220 1225	
GAA ACC GCA CTA GAA GGT GGA GAC AAG GGG CAT TAT GCC GGC CAC GAG	3868
Glu Thr Ala Leu Gly Gly Gly Asp Lys Gly His Tyr Ala Gly His Glu	
1230 1235 1240	
ATT GTG AGG TAT GGA AGT GGC CCA GCA CTG TCA ACT AAA ACA AAA TTC	3916
Ile Val Arg Tyr Gly Ser Gly Pro Ala Leu Ser Thr Lys Thr Lys Phe	
1245 1250 1255	

Table 3, continued

TGG AGG TCC TCC CCA GAA CCA CTG CCC CCC GGA GTA TAT GAG CCA GCA	3964
Trp Arg Ser Ser Pro Glu Pro Leu Pro Pro Gly Val Tyr Glu Pro Ala	
1260 1265 1270	
TAC CTG GGG GGC AAG GAC CCC CGT GTA CAG AAT GGC CCA TCC CTA CAA	4012
Tyr Leu Gly Gly Lys Asp Pro Arg Val Gln Asn Gly Pro Ser Leu Gln	
1275 1280 1285	
CAG GTA CTA CGT GAC CAA CTG AAA CCC TTT GCG GAC CCC CGC GGC CGC	4060
Gln Val Leu Arg Asp Gln Leu Lys Pro Phe Ala Asp Pro Arg Gly Arg	
1290 1295 1300 1305	
ATG CCT GAG CCT GGC CTA CTG GAG GCT GCG GTT GAG ACT GTA ACA TCC	4108
Met Pro Glu Pro Gly Leu Leu Glu Ala Val Glu Thr Val Thr Ser	
1310 1315 1320	
ATG TTA GAA CAG ACA ATG GAT ACC CCA AGC CCG TGG TCT TAC GCT GAT	4156
Met Leu Glu Gln Thr Met Asp Thr Pro Ser Pro Trp Ser Tyr Ala Asp	
1325 1330 1335	
GCC TGC CAA TCT CTT GAC AAA ACT ACT AGT TCG GGG TAC CCT CAC CAT	4204
Ala Cys Gln Ser Leu Asp Lys Thr Thr Ser Ser Gly Tyr Pro His His	
1340 1345 1350	
AAA AGG AAG AAT GAT GAT TGG AAT GGC ACC ACC TTC GTT GGA GAG CTC	4252
Lys Arg Lys Asn Asp Asp Trp Asn Gly Thr Thr Phe Val Gly Glu Leu	
1355 1360 1365	
GGT GAG CAA GCT GCA CAC GCC AAC AAT ATG TAT GAG AAT GCT AAA CAT	4300
Gly Glu Gln Ala Ala His Ala Asn Asn Met Tyr Glu Asn Ala Lys His	
1370 1375 1380 1385	
ATG AAA CCC ATT TAC ACT GCA GCC TTA AAA GAT GAA CTA GTC AAG CCA	4348
Met Lys Pro Ile Tyr Thr Ala Ala Leu Lys Asp Glu Leu Val Lys Pro	
1390 1395 1400	
GAA AAG ATT TAT CAA AAA GTC AAG AAG CGT CTA CTA TGG GGC GCC GAT	4396
Glu Lys Ile Tyr Gln Lys Val Lys Lys Arg Leu Leu Trp Gly Ala Asp	
1405 1410 1415	
CTC GGA ACA GTG GTC AGG GCC GCC CGG GCT TTT GGC CCA TTT TGT GAC	4444
Leu Gly Thr Val Val Arg Ala Ala Arg Ala Phe Gly Pro Phe Cys Asp	
1420 1425 1430	
GCT ATA AAA TCA CAT GTC ATC AAA TTG CCA ATA AAA GTT GGC ATG AAC	4492
Ala Ile Lys Ser His Val Ile Lys Leu Pro Ile Lys Val Gly Met Asn	
1435 1440 1445	
ACA ATA GAA GAT GGC CCC CTC ATC TAT GCT GAG CAT GCT AAA TAT AAG	4540
Thr Ile Glu Asp Gly Pro Leu Ile Tyr Ala Glu His Ala Lys Tyr Lys	
1450 1455 1460 1465	
AAT CAT TTT GAT GCA GAT TAT ACA GCA TGG GAC TCA ACA CAA AAT AGA	4588
Asn His Phe Asp Ala Asp Tyr Thr Ala Trp Asp Ser Thr Gln Asn Arg	
1470 1475 1480	
CAA ATT ATG ACA GAA TCC TTC TCC ATT ATG TCG CGC CTT ACG GCC TCA	4636
Gln Ile Met Thr Glu Ser Phe Ser Ile Met Ser Arg Leu Thr Ala Ser	
1485 1490 1495	
CCA GAA TTG GCC GAG GTT GTG GCC CAA GAT TTG CTA GCA CCA TCT GAG	4684
Pro Glu Leu Ala Glu Val Val Ala Gln Asp Leu Leu Ala Pro Ser Glu	
1500 1505 1510	

Table 3, continued

ATG GAT GTA GGT GAT TAT GTC ATC AGG GTC AAA GAG GGG CTG CCA TCT Met Asp Val Gly Asp Tyr Val Ile Arg Val Lys Glu Gly Leu Pro Ser 1515 1520 1525	4732
GGA TTC CCA TGT ACT TCC CAG GTG AAC AGC ATA AAT CAC TGG ATA ATT Gly Phe Pro Cys Thr Ser Gln Val Asn Ser Ile Asn His Trp Ile Ile 1530 1535 1540 1545	4780
ACT CTC TGT GCA CTG TCT GAG GCC ACT GGT TTA TCA CCT GAT GTG GTG Thr Leu Cys Ala Leu Ser Glu Ala Thr Gly Leu Ser Pro Asp Val Val 1550 1555 1560	4828
CAA TCC ATG TCA TAT TTC TCA TTT TAT GGT GAT GAT GAG ATT GTG TCA Gln Ser Met Tyr Phe Ser Phe Tyr Gly Asp Asp Glu Ile Val Ser 1565 1570 1575	4876
ACT GAC ATA GAT TTT GAC CCA GCC CGC CTC ACT CAA ATT CTC AAG GAA Thr Asp Ile Asp Phe Asp Pro Ala Arg Leu Thr Gln Ile Leu Lys Glu 1580 1585 1590	4924
TAT GGC CTC AAA CCA ACA AGG CCT GAC AAA ACA GAA GGA CCA ATA CAA Tyr Gly Leu Lys Pro Thr Arg Pro Asp Lys Thr Glu Gly Pro Ile Gln 1595 1600 1605	4972
GTG AGG AAA AAT GTG GAT GGA CTG GTC TTC TTG CGG CGC ACC ATT TCC Val Arg Lys Asn Val Asp Gly Leu Val Phe Leu Arg Arg Thr Ile Ser 1610 1615 1620 1625	5020
CGT GAT GCG GCA GGG TTC CAA GGC AGG TTA GAT AGG GCT TCG ATT GAA Arg Asp Ala Ala Gly Phe Gln Gly Arg Leu Asp Arg Ala Ser Ile Glu 1630 1635 1640	5068
CGC CAA ATC TTC TGG ACC CGC GGG CCC AAT CAT TCA GAT CCA TCA GAG Arg Gln Ile Phe Trp Thr Arg Gly Pro Asn His Ser Asp Pro Ser Glu 1645 1650 1655	5116
ACT CTA GTG CCA CAC ACT CAA AGA AAA ATA CAG TTG ATT TCA CTT CTA Thr Leu Val Pro His Thr Gln Arg Lys Ile Gln Leu Ile Ser Leu Leu 1660 1665 1670	5164
GGG GAA GCT TCA CTC CAT GGT GAG AAA TTT TAC AGA AAG ATT TCC AGC Gly Glu Ala Ser Leu His Gly Glu Lys Phe Tyr Arg Lys Ile Ser Ser 1675 1680 1685	5212
AAG GTC ATA CAT GAA ATC AAG ACT GGT GGA TTG GAA ATG TAT GTC CCA Lys Val Ile His Glu Ile Lys Thr Gly Gly Leu Glu Met Tyr Val Pro 1690 1695 1700 1705	5260
GGA TGG CAG GCC ATG TTC CGC TGG ATG CGC TTC CAT GAC CTC GGA TTG Gly Trp Gln Ala Met Phe Arg Trp Met Arg Phe His Asp Leu Gly Leu 1710 1715 1720	5308
TGG ACA GGA GAT CGC GAT CTT CTG CCC GAA TTC GTA AAT GAT GAT GGC Trp Thr Gly Asp Arg Asp Leu Leu Pro Glu Phe Val Asn Asp Asp Gly 1725 1730 1735	5356
GTC TAAGGACGCT ACATCAAGCG TGGATGGCGC TAGTGGCGCT GGTCAGTTGG Val	5409

Table 4. The amino acid sequence deduced from nucleotides 5346 through 6935 of the Norwalk virus genome shown in Table 2.

CGTAA ATG ATG ATG GCG TCT AAG GAC GCT ACA TCA AGC GTG GAT GGC Met Met Met Ala Ser Lys Asp Ala Thr Ser Ser Val Asp Gly 1 5 10	5387
GCT AGT GGC GCT GGT CAG TTG GTA CCG GAG GTT AAT GCT TCT GAC CCT Ala Ser Gly Ala Gly Gln Leu Val Pro Glu Val Asn Ala Ser Asp Pro 15 20 25 30	5435
CTT GCA ATG GAT CCT GTA GCA GGT TCT TCG ACA GCA GTC GCG ACT GCT Leu Ala Met Asp Pro Val Ala Gly Ser Ser Thr Ala Val Ala Thr Ala 35 40 45	5483
GGA CAA GTT AAT CCT ATT GAT CCC TGG ATA ATT AAT AAT TTT GTG CAA Gly Gln Val Asn Pro Ile Asp Pro Trp Ile Ile Asn Asn Phe Val Gln 50 55 60	5531
GCC CCC CAA GGT GAA TTT ACT ATT TCC CCA AAT AAT ACC CCC GGT GAT Ala Pro Gln Gly Glu Phe Thr Ile Ser Pro Asn Asn Thr Pro Gly Asp 65 70 75	5579
GTT TTG TTT GAT TTG AGT TTG GGT CCC CAT CTT AAT CCT TTC TTG CTC Val Leu Phe Asp Leu Ser Leu Gly Pro His Leu Asn Pro Phe Leu Leu 80 85 90	5627
CAT CTA TCA CAA ATG TAT AAT GGT TGG GTT GGT AAC ATG AGA GTC AGG His Leu Ser Gln Met Tyr Asn Gly Trp Val Gly Asn Met Arg Val Arg 95 100 105 110	5675
ATT ATG CTA GCT GGT AAT GCC TTT ACT GCG GGG AAG ATA ATA GTT TCC Ile Met Leu Ala Gly Asn Ala Phe Thr Ala Gly Lys Ile Ile Val Ser 115 120 125	5723
TGC ATA CCC CCT GGT TTT GGT TCA CAT AAT CTT ACT ATA GCA CAA GCA Cys Ile Pro Pro Gly Phe Gly Ser His Asn Leu Thr Ile Ala Gln Ala 130 135 140	5771
ACT CTC TTT CCA CAT GTG ATT GCT GAT GTT AGG ACT CTA GAC CCC ATT Thr Leu Phe Pro His Val Ile Ala Asp Val Arg Thr Leu Asp Pro Ile 145 150 155	5819
GAG GTG CCT TTG GAA GAT GTT AGG AAT GTT CTC TTT CAT AAT AAT GAT Glu Val Pro Leu Glu Asp Val Arg Asn Val Leu Phe His Asn Asn Asp 160 165 170	5867
AGA AAT CAA CAA ACC ATG CGC CTT GTG TGC ATG CTG TAC ACC CCC CTC Arg Asn Gln Gln Thr Met Arg Leu Val Cys Met Leu Tyr Thr Pro Leu 175 180 185 190	5915
CGC ACT GGT GGT GGT ACT GGT GAT TCT TTT GTA GTT GCA GGG CGA GTT Arg Thr Gly Gly Gly Thr Gly Asp Ser Phe Val Val Ala Gly Arg Val 195 200 205	5963
ATG ACT TGC CCC AGT CCT GAT TTT AAT TTC TTG TTT TTA GTC CCT CCT Met Thr Cys Pro Ser Pro Asp Phe Asn Phe Leu Phe Leu Val Pro Pro 210 215 220	6011
ACG GTG GAG CAG AAA ACC AGG CCC TTC ACA CTC CCA AAT CTG CCA TTG Thr Val Glu Gln Lys Thr Arg Pro Phe Thr Leu Pro Asn Leu Pro Leu 225 230 235	6059

Table 4, continued

AGT TCT CTG TCT AAC TCA CGT GCC CCT CTC CCA ATC AGT AGT ATG GGC Ser Ser Leu Ser Asn Ser Arg Ala Pro Leu Pro Ile Ser Ser Met Gly 240 245 250	6107
ATT TCC CCA GAC AAT GTC CAG AGT GTG CAG TTC CAA AAT GGT CGG TGT Ile Ser Pro Asp Asn Val Gln Ser Val Gln Phe Gln Asn Gly Arg Cys 255 260 265 270	6155
ACT CTG GAT GGC CGC CTG GTT GGC ACC ACC CCA GTT TCA TTG TCA CAT Thr Leu Asp Gly Arg Leu Val Gly Thr Thr Pro Val Ser Leu Ser His 275 280 285	6203
GTT GCC AAG ATA AGA GGG ACC TCC AAT GGC ACT GTA ATC AAC CTT ACT Val Ala Lys Ile Arg Gly Thr Ser Asn Gly Thr Val Ile Asn Leu Thr 290 295 300	6251
GAA TTG GAT GGC ACA CCC TTT CAC CCT TTT GAG GGC CCT GCC CCC ATT Glu Leu Asp Gly Thr Pro Phe His Pro Phe Glu Gly Pro Ala Pro Ile 305 310 315	6299
GGG TTT CCA GAC CTC GGT GGT TGT GAT TGG CAT ATC AAT ATG ACA CAG Gly Phe Pro Asp Leu Gly Gly Cys Asp Trp His Ile Asn Met Thr Gln 320 325 330	6347
TTT GGC CAT TCT AGC CAG ACC CAG TAT GAT GTA GAC ACC ACC CCT GAC Phe Gly His Ser Ser Gln Thr Gln Tyr Asp Val Asp Thr Thr Pro Asp 335 340 345 350	6395
ACT TTT GTC CCC CAT CTT GGT TCA ATT CAG GCA AAT GGC ATT GGC AGT Thr Phe Val Pro His Leu Gly Ser Ile Gln Ala Asn Gly Ile Gly Ser 355 360 365	6443
GGT AAT TAT GTT GGT GTT CTT AGC TGG ATT TCC CCC CCA TCA CAC CCG Gly Asn Tyr Val Gly Val Leu Ser Trp Ile Ser Pro Pro Ser His Pro 370 375 380	6491
TCT GGC TCC CAA GTT GAC CTT TGG AAG ATC CCC AAT TAT GGG TCA AGT Ser Gly Ser Gln Val Asp Leu Trp Lys Ile Pro Asn Tyr Gly Ser Ser 385 390 395	6539
ATT ACG GAG GCA ACA CAT CTA GCC CCT TCT GTA TAC CCC CCT GGT TTC Ile Thr Glu Ala Thr His Leu Ala Pro Ser Val Tyr Pro Pro Gly Phe 400 405 410	6587
GGA GAG GTA TTG GTC TTT TTC ATG TCA AAA ATG CCA GGT CCT GGT GCT Gly Glu Val Leu Val Phe Phe Met Ser Lys Met Pro Gly Pro Gly Ala 415 420 425 430	6635
TAT AAT TTG CCC TGT CTA TTA CCA CAA GAG TAC ATT TCA CAT CTT GCT Tyr Asn Leu Pro Cys Leu Leu Pro Gln Glu Tyr Ile Ser His Leu Ala 435 440 445	6683
AGT GAA CAA GCC CCT ACT GTA GGT GAG GCT GCC CTG CTC CAC TAT GTT Ser Glu Gln Ala Pro Thr Val Gly Glu Ala Ala Leu Leu His Tyr Val 450 455 460	6731
GAC CCT GAT ACC GGT CGG AAT CTT GGG GAA TTC AAA GCA TAC CCT GAT Asp Pro Asp Thr Gly Arg Asn Leu Gly Glu Phe Lys Ala Tyr Pro Asp 465 470 475	6779
GGT TTC CTC ACT TGT GTC CCC AAT GGG GCT AGC TCG GGT CCA CAA CAG Gly Phe Leu Thr Cys Val Pro Asn Gly Ala Ser Ser Gly Pro Gln Gln 480 485 490	6827

Table 4, continued

[illegible]

Table 5. The amino acid sequence deduced from nucleotides 6938 through 7573 of the Norwalk virus genome shown in Table 2.

CCAGCTCGGC AAGAGGTAGG CTTGGTCTGC GCCGATA ATG GCC CAA GCC ATA ATT	6955
Met Ala Gln Ala Ile Ile	
1 5	
GGT GCA ATT GCT GCT TCC ACA GCA GGT AGT GCT CTG GGA GCG GGC ATA	7003
Gly Ala Ile Ala Ala Ser Thr Ala Gly Ser Ala Leu Gly Ala Gly Ile	
10 15 20	
CAG GTT GGT GGC GAA GCG GCC CTC CAA AGC CAA AGG TAT CAA CAA AAT	7051
Gln Val Gly Gly Glu Ala Ala Leu Gln Ser Gln Arg Tyr Gln Gln Asn	
25 30 35	
TTG CAA CTG CAA GAA AAT TCT TTT AAA CAT GAC AGG GAA ATG ATT GGG	7099
Leu Gln Leu Gln Glu Asn Ser Phe Lys His Asp Arg Glu Met Ile Gly	
40 45 50	
TAT CAG GTT GAA GCT TCA AAT CAA TTA TTG GCT AAA AAT TTG GCA ACT	7147
Tyr Gln Val Glu Ala Ser Asn Gln Leu Leu Ala Lys Asn Leu Ala Thr	
55 60 65 70	
AGA TAT TCA CTC CTC CGT GCT GGG GGT TTG ACC AGT GCT GAT GCA GCA	7195
Arg Tyr Ser Leu Leu Arg Ala Gly Gly Leu Thr Ser Ala Asp Ala Ala	
75 80 85	
AGA TCT GTG GCA GGA GCT CCA GTC ACC CGC ATT GTA GAT TGG AAT GGC	7243
Arg Ser Val Ala Gly Ala Pro Val Thr Arg Ile Val Asp Trp Asn Gly	
90 95 100	
GTG AGA GTG TCT GCT CCC GAG TCC TCT GCT ACC ACA TTG AGA TCC GGT	7291
Val Arg Val Ser Ala Pro Glu Ser Ser Ala Thr Thr Leu Arg Ser Gly	
105 110 115	
GGC TTC ATG TCA GTT CCC ATA CCA TTT GCC TCT AAG CAA AAA CAG GTT	7339
Gly Phe Met Ser Val Pro Ile Pro Phe Ala Ser Lys Gln Lys Gln Val	
120 125 130	
CAA TCA TCT GGT ATT AGT AAT CCA AAT TAT TCC CCT TCA TCC ATT TCT	7387
Gln Ser Ser Gly Ile Ser Asn Pro Asn Tyr Ser Pro Ser Ser Ile Ser	
135 140 145 150	
CGA ACC ACT AGT TGG GTC GAG TCA CAA AAC TCA TCG AGA TTT GGA AAT	7435
Arg Thr Thr Ser Trp Val Glu Ser Gln Asn Ser Ser Arg Phe Gly Asn	
155 160 165	
CTT TCT CCA TAC CAC GCG GAG GCT CTC AAT ACA GTG TGG TTG ACT CCA	7483
Leu Ser Pro Tyr His Ala Glu Ala Leu Asn Thr Val Trp Leu Thr Pro	
170 175 180	
CCC GGT TCA ACA GCC TCT TCT ACA CTG TCT TCT GTG CCA CGT GGT TAT	7531
Pro Gly Ser Thr Ala Ser Ser Thr Leu Ser Ser Val Pro Arg Gly Tyr	
185 190 195	
TTC AAT ACA GAC AGG TTG CCA TTA TTC GCA AAT AAT AGG CGA	7573
Phe Asn Thr Asp Arg Leu Pro Leu Phe Ala Asn Asn Arg Arg	
200 205 210	

Table 6. Primers used for detection of Norwalk-related virus by PCR

P-35	4944	5'	CTT GTT GGT TTG AGG CCA TAT	4924
P-36	4475	5'	ATA AAA GTT GGC ATG AAC A	4493
P-39	4878	5'	GTT GAC ACA ATC TCA TCA TC	4859
P-69	4721	5'	GGC CTG CCA TCT GGA TTG CC	4740
P-78	1670	5'	GGG CCC CCT GGT ATA GGT AA	1689
P-80	1931	5'	TGG TGA TGA CTA TAG CAT CAG ACA CAA A	1958
P-56*	4903	5'	ACT CAC CCA AAT CCT CCA	4920
P-23	5230	5'	GTT CTG ACC ACC TAA CCT	5247
P-42	5595	5'	AGT TTG GGT CCC CAT CTT AAT CCT TT	5620
P-55	5730	5'	TGA ACC AAA ACC AGG GGG	5747
P-58*	5210	5'	AGC AAA GTC ATA CAT GAA AT	5229
P-59	5634	5'	CCA TTA TAC ATT TGT AG	5650
P-60*	5712	5'	ATT ATA GTT TCT TGC ATA	5729
P-61	6115	5'	CAC ACT CTG GAC ATT GTC TG	6134
P-72	6296	5'	CAT TGG GTT TCC AGA CCT A	6313
P-63	6511	5'	ATA ATT GGG GAT CTT CCA AA	6530
P-76*	6095	5'	TAG TGG CAT GGG TAT TTC	6114
P-77	6316	5'	TAT GCC AAT CAC AGC CAC	6333
P-64	6491	5'	GTC TGG CTC CCA AGT TGA CC	6510
P-75	6726	5'	CGG TAT CAG GGT CAA CAT	6744
P-74	6707	5'	TGA GGC TGC CCT GCT CCA	6724
P-3	7009	5'	CCA CCG CTG TCC GGG AGG	7027
P-36 (New) #	4475	5'	GTT GCT GTT GGC ATT AAC A	4493

*Based on KY89 sequence

Based HuVc Sapporo sequence

Table 7. Detection of Norwalk and Norwalk-Related Viruses

Virus	NV ELISA (36-35)	Polymerase		2C PCR (78-80)	Size of cDNA (bp)	YGDD motif	% na similarity to NV	% aa similarity to NV	Virus or Reference
		PCR (36-35)	PCR (69-39)						
8FIIa NV/68	+	+	+	+	470	yes	100	100	A. Kapikian
SRSV-3/88	+	+	ND	ND	470	yes	87	99	T. Ando
SRSV/KY/89	+	+	+	+	470	yes	87	99	I. Oishi
SRSV/CDC6/91	+	-	+	ND	118*	N/A	80	89	C. Moe/R. Glass
Desert Storm/90	-	+	ND	ND	406*	yes	73	85	K. Green/J. Lew
SRSV/UT/88	-	+	+	ND	118*	N/A	71	82	P. Johnson
SMA/79	-	+	ND	ND	464	yes	63	60	P. Madore
SRSV/Cambridge, UK/92	-	+	+	-	469	yes	63	60	U. Desselberger
Toronto/91	ND	+	ND	ND	470	yes	63	60	K. Green/M. Petric
CDC 32	-	-	+	ND	117*	N/A	62	66	C. Moe/R. Glass
HuCV Sapporo/82	-	+	ND	ND	488	yes	26	24	S. Nakata/D. Matson
Primate CV	ND	+	ND	ND	~470	yes	20	ND#	A. Smith/D. Matson
HuCV Houston DCC	ND	+	ND	ND	485	yes	17	12	D. Matson
HuCV Houston Child	ND	+	ND	ND	474	yes	9.8	6	D. Matson
Astro/CDC 37	-	+	+	ND	400	N/A	0	0	C. Moe/R. Glass

ND = Not done. N/A = Not available

* Internal primers were used to amplify this agent.

† The primate CV 35-36 PCR product sequence is not yet complete. Similarity information is based upon the partial sequence.

° Size of 69-39 PCR cDNA product. Primers 69-39 are located inside 36-35 on the Norwalk virus genome sequence. All others are the sizes of the cDNA products made using primers 36-35.

Table 8. Nucleotide Homologies of Different Caliciviruses in the Primer 36 Region.

Virus Strain	Nucleotide*																		
	A	T	A	A	A	A	G	T	T	G	G	C	A	T	G	A	A	C	A
Norwalk Virus	A	T	A	A	A	A	G	T	T	G	G	C	A	T	G	A	A	C	A
HuCV Sapporo	G	-	T	G	C	T	-	-	-	-	-	-	-	-	-	-	T	G	G
Feline CV F4	-	T	T	C	C	T	-	-	-	-	-	-	-	-	-	-	T	G	G
Feline CV CFI	-	-	-	C	C	G	-	-	-	-	-	-	-	-	-	-	T	G	G
Feline CV F9	-	-	-	C	-	G	-	-	-	-	-	T	-	-	-	-	T	G	G
Rabbit CV	-	-	T	G	C	-	-	-	-	-	-	G	-	-	-	T	G	-	C

* "-" means nucleotide is identical at that site for the strain. A new letter at a site indicates the nucleotide differs at that site. Primer 36 extends across the first 19 nucleotides in the Norwalk sequence above. Primer "new 36" is the first 19 nucleotides of the HuCV Sapporo sequence.

Table 9. Characterization of Serum from Animals Immunized Parenterally with Recombinant Norwalk Virus Particles^a

ELISA Titer of Indicated Serum with Norwalk Virus Particles

<u>Species Immunized</u>	<u>Pre-immune</u>	<u>Post-immune^c</u>
6 mice	<100 ^b	>10 ⁶
4 guinea pigs	<100	>10 ⁶
2 rabbits	<100	>10 ⁶

-
- a Two to six animals of each species were immunized with purified recombinant Norwalk virus protein produced using the baculovirus expression system. Serum was collected before or after three immunizations with antigen (80mg for mice, 200mg for guinea pigs and 300mg for rabbits) and tested for reactivity with antigen coated on ELISA plates.
- b Lowest dilution tested was 1:100.
- c These sera also were used as capture and detector antibodies to establish an ELISA to detect Norwalk virus antigen.

Table 10. Nucleotide and amino acid sequence of human calicivirus Houston cDNA:

GG CCA TGT TAT AGT GGT GTT CAC ATG AAA GAT GGC GAC AAG ATG P C Y S G V H M K K D G D K M	44	HuCV Houston HuCV Houston amino acid
TTG ATA GAT GCC AAT CTT CCT TAC AAC CAG AAA TTA ACT ACT ATG L I D A N L P Y N Q K K L T T M	89	HuCV Houston HuCV Houston amino acid
ATT CAT GAG ACT AGG CAT AGG ATA GGA CAG TAT ATA GAT AAT ACT I H E T R H R I G Q Y I D N T	134	HuCV Houston HuCV Houston amino acid
TTT GGA AAG ACA TTT AGA CAT GGA TTG ACA AAA CCT GCT GAC AAG F G K T F R H G L T K P A D K	179	HuCV Houston HuCV Houston amino acid
ACT GTA GAT TTG ATC TAT AAG ACA TTG AAT TAT GAT GAT TTT CTG T V D L I Y K T L N Y D D F L	224	HuCV Houston HuCV Houston amino acid
GCA ATA ATG CTA ATC ATA TAT GGG CAA AAG TCG GCC ACT AAT ACG A I M L I I Y G Q K S A T N T	269	HuCV Houston HuCV Houston amino acid
GAG TTG CAA TTC TTG ATG GAG AAA CTT AGA GGT TAT GAA TCT ACA E L Q F L M E K L R G Y E S T	314	HuCV Houston HuCV Houston amino acid
ATG GAT GAC ATA GGG AAA GTC TAT GGA GAT GAT AAA ATG AGA GAT M D D I G K V Y G D D K M R D	359	HuCV Houston HuCV Houston amino acid
ATA ATC AAG AAT ATT TCT GAT GAT GAC ATA AAG AGT CTT TTA GGG I I K N I S D D I K S L L G	404	HuCV Houston HuCV Houston amino acid
GAG ATA AAT AGT GAT TAT TCT GGT AAG NAT E I N S D Y S G K X	434	HuCV Houston Houston amino acid

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Matson, David O
Estes, Mary K
Jiang, Xi
Graham, David Y
- (ii) TITLE OF INVENTION: Methods and Reagents to Detect and Characterize Norwalk and Related Viruses
- (iii) NUMBER OF SEQUENCES: 75
- (iv) CORRESPONDENCE ADDRESS:
 - (A) ADDRESSEE: Fulbright & Jaworski Patent Dept
 - (B) STREET: 1301 McKinney, Suite 5100
 - (C) CITY: Houston
 - (D) STATE: Texas
 - (E) COUNTRY: USA
 - (F) ZIP: 77010-3095
- (v) COMPUTER READABLE FORM:
 - (A) MEDIUM TYPE: Floppy disk
 - (B) COMPUTER: IBM PC compatible
 - (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - (D) SOFTWARE: PatentIn Release #1.0, Version #1.25
- (vi) CURRENT APPLICATION DATA:
 - (A) APPLICATION NUMBER:
 - (B) FILING DATE:
 - (C) CLASSIFICATION:
- (viii) ATTORNEY/AGENT INFORMATION:
 - (A) NAME: Launer, Charlene A
 - (B) REGISTRATION NUMBER: 33,035
 - (C) REFERENCE/DOCKET NUMBER: D-5526
- (ix) TELECOMMUNICATION INFORMATION:
 - (A) TELEPHONE: 713-651-3634
 - (B) TELEFAX: 713-651-5246
 - (C) TELEX: Western Union 762829

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 7753 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: cDNA
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: Norwalk virus
 - (B) STRAIN: 8FIIa
 - (C) INDIVIDUAL ISOLATE: 8FIIa
- (vii) IMMEDIATE SOURCE:
 - (B) CLONE: pUCNV-953 and its derivatives
- (ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 146..5359
(D) OTHER INFORMATION: /note= "The protein encoded by nucleotides 146 through 5359 is eventually cleaved to make at least a picornavirus 2c-like protein, a 3C-like protease and an RNA-dependent RNA polymerase."

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 5346..6935
(D) OTHER INFORMATION: /note= "Nucleotides 5346 through 5359 are used for coding two different amino acid sequences: the first is the amino acid is coded by nucleotide 146 through 5359, the second by nucleotides 5346 through 6935."

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 6938..7573

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

GGCGTCAAAA	GACGTCGTTC	CTACTGCTGC	TAGCAGTGAA	AATGCTAACA	ACAATAGTAG	60
TATTAAGTCT	CGTCTATTGG	CGAGACTCAA	GGGTTCAGGT	GGGGCTACGT	CCCCACCCAA	120
CTCGATAAAG	ATAACCAACC	AAGATATGGC	TCCTGGGGCTG	ATTGGACAGG	TCCCAGCGCC	180
AAAGGCCACA	TCCGTCGATG	TCCCTAAACA	ACAGAGGGAT	AGACCACCAC	GGACTGTTGC	240
CGAAGTTCAA	CAAAATTTCG	GTTGGACTGA	GAGACCACAA	GACCAGAATG	TTAAGACGTG	300
GGATGAGCTT	GACCACACAA	CAAAACAACA	GATACTTGAT	GAACACGCTG	AGTGGTTTGA	360
TGCCGGTGGC	TTAGGTCCAA	GTACACTACC	CACTAGTCAT	GAACGGTACA	CACATGAGAA	420
TGATGAAGGC	CACCAGGTAA	AGTGGTCGGC	TAGGGAAGGT	GTAGACCTTG	GCATATCCGG	480
GCTCACGACG	GTGTCTGGGC	CTGAGTGGA	TATGTGCCCG	CTACCACCAG	TTGACCAAAG	540
GAGCACGACA	CCTGCCAACTG	AGCCCACAAT	TGGTGACATG	ATCGAATTCT	ATGAAGGGCA	600
CATCTATCAT	TATGCTATAT	ACATAGGTCA	AGGCAAGACG	GTGGGTGTAC	ACTCCCCTCA	660
AGCAGCCTTC	TCAATATCGA	GGATCACCAT	ACAGCCCAT	TCAGCTTGGT	GGCGAGTCTG	720
TTATGTCCCA	CAACCAAAAC	AGAGGCTCAC	ATACGACCAA	CTCAAAGAAT	TAGAAAATGA	780
ACCATGGCCG	TATGCCCGCAG	TCACGAACAA	CTGCTTCGAA	TTTTGTTGCC	AGGTCATGTG	840
CTTGGAAGAT	ACTTGGTTGC	AAAGGAAGCT	CATCTCCTCT	GGCCGGTTTT	ACCACCCGAC	900
CCAAGATTGG	TCCCGAGACA	CTCCAGAATT	CCAACAAGAC	AGCAAGTTAG	AGATGGTTAG	960
GGATGTCAGT	CTAGCCGCTA	TAAATGGGTT	GGTGTCCGGG	CCATTAAAG	ATCTTCTGGG	1020
TAAGCTCAAA	CCCTTGAAACG	TGCTTAACTT	ACTTTCAAAC	TGTGATTGGA	CGTTCATGGG	1080
GGTCTGTGGAG	ATGGTGTGCC	TCCTTTTAGA	ACTCTTTGGA	ATCTTTTGGG	ACCCACCTGA	1140
TGTTTCCAAC	TTTATAGCTT	CACCTCTGCC	AGATTTCAT	CTACAGGGCC	CCGAGGACCT	1200
TGCCAGGGAT	CTCGTGCCAA	TAGTATTGGG	GGGGATCGGC	TTAGCCATAG	GATTCAACAG	1260
AGACAAGGTA	AGTAAGATGA	TGAAGAATGC	TGTTGATGGA	CTTCGTGCGG	CAACCCAGCT	1320
CGGTCAATAT	GGCCTAGAAA	TATTCTCATT	ACTAAAGAAG	TACTTCTTCG	GTGGTGATCA	1380
AACAGAGAAA	ACCCTAAAAA	ATATTGAGTC	AGCAGTTATA	GATATGGAAG	TACTATCATC	1440
TACATCAGTG	ACTCAGCTCG	TGAGGGACAA	ACAGTCTGCA	CGGGCTTATA	TGGCCATCTT	1500
AGATAATGAA	GAAGAAAAGG	CAAGGAAATT	ATCTGTCAGG	AATGCCGACC	CACACGTAGT	1560
ATCCTCTACC	AATGCTCTCA	TATCCCGGAT	CTCAATGGCT	AGGGCTGCAT	TGGCCAAGGC	1620
TCAAGCTGAA	ATGACCAGCA	GGATGCGTCC	TGTGGTCATT	ATGATGTGTG	GGCCCCCTGG	1680
TATAGGTAAA	ACCAAGGCAG	CAGAACATCT	GGCTAAACGC	CTAGCCAATG	AGATACGGCC	1740
TGGTGGTAAG	GTTGGGCTGG	TCCCACGGGA	GGCAGTGGAT	CATTGGGATG	GATATCACGG	1800
AGAGGAAGTG	ATGCTGTGGG	ACGACTATGG	AATGACAAAG	ATACAGGAAG	ACTGTAATAA	1860
ACTGCAAGCC	ATAGCCGACT	CAGCCCCCCT	AACACTCAAT	TGTGACCGAA	TAGAAAACAA	1920
GGGAATGCAA	TTTGTGCTG	ATGCTATAGT	CATCACCACC	AATGCTCCTG	GCCCAGCCCC	1980
AGTGGACTTT	GTCAACCTCG	GGCCTGTTTG	CCGAAGGGTG	GACTTCCTTG	TGTATTGCAC	2040
GGCACCTGAA	GTTGAACACA	CGAGGAAAGT	CAGTCCTGGG	GACACAACCTG	CAGTGAAAGA	2100
CTGCTTCAAG	CCCGATTCT	CACATCTAAA	AATGGAGTTG	GCTCCCCAAG	GGGGCTTTGA	2160
TAACORAGGG	AATACCCCGT	TTGGTAAGGG	TGTGATGAAG	CCCACCCCA	TAAACAGGCT	2220
GTTAATCCAG	GCTGTAGCCT	TGACGATGGA	GAGACAGGAT	GAGTTCCAAC	TCCAGGGGCC	2280
TACGTATGAC	TTTGATACTG	ACAGAGTAGC	TGCGTTACAG	AGGATGGCCC	GAGCCAACGG	2340
GTTGGGCTCT	ATATCCATGG	CCTCCCTAGG	CAAAAAGCTA	CGCAGTGTC	CCACTATTGA	2400
AGGATTAAAG	AATGCTCTAT	CAGGCTATAA	ATATCAAAA	TGCAGTATAC	AATGGCAGTC	2460
AAGGGTGTAC	ATTATAGAA	CAGATGCTGC	CAGTGTACAA	ATCAAGAAG	ACAAGCAAGC	2520
TTTGACCCCT	CTGCAGCAGA	CAATTACAC	GGCCTCACTT	GCCATCACTC	GACTCAAAGC	2580

AGCTAGGGCT	GTGGCATAAG	CTTCATGTTT	CCAGTCCGCC	ATAACTACCA	TACTACAAAT	2640
GGCGGGATCT	GCGCTCGTTA	TTAATCGAGC	GGTCAAGCGT	ATGTTTGGTA	CCCGTACAGC	2700
AGCCATGGCA	TTAGAAGGAC	CTGGGAAAGA	ACATAATTGC	AGGGTCCATA	AGGCTAAGGA	2760
AGCTGGAAG	GGGCCCATAG	GTCATGATGA	CATGGTAGAA	AGGTTTGGCC	TATGTGAAAC	2820
TGAAGAGGAG	GAGAGTGAGG	ACCAAATTCA	AATGGTACCA	AGTGATGCCG	TCCCAGAAAG	2880
AAAGAACAAA	GGCAAGACCA	AAAAGGGACG	TGGTCGCAAA	AATAACTATA	ATGCATTCTC	2940
TCGCCGTGGT	CTGAGTGATG	AAGAATATGA	AGAGTACAAA	AAGATCAGAG	AAGAAAAGAA	3000
TGGCAATTAT	AGTATACAAG	AATACTTGGA	GGACCGCCAA	CGATATGAGG	AAGAATTAGC	3060
AGAGGTACAG	GCAGGTGGTG	ATGGTGGCAT	AGGAGAAACT	GAAATGAAA	TCCGTACACAG	3120
GGTCTTCTAT	AAATCCAAGA	GTAAGAAACA	CCAACAAGAG	CAACGGCGAC	AACCTGGTCT	3180
AGTGACTGGA	TCAGACATCA	GAAAACGTAA	GCCCATTGAC	TGGACCCCGC	CAAAGAATGA	3240
ATGGGCAGAT	GCAGACAGAG	AGGTGGATTA	TGAAGAAAAG	ATCAATTTTG	AAGCTCCCCC	3300
GACACTATGG	AGCCGAGTCA	CAAAGTTTGG	ATCAGGATGG	GGCTTTTGGG	TCAGCCCGAC	3360
AGTGTTTCATC	ACAACACAC	ATGTAGTGCC	AACTGGTGTG	AAAGAATTCT	TTGGTGAGCC	3420
CCTATCTAGT	ATAGCAATCC	ACCAAGCAGG	TGAGTTTACA	CAATTCAGGT	TCTCAAAGAA	3480
AATGCCGCTT	AGTATGACAG	GTATGGTCTT	TGAAGAAAGT	TGCCCTGAAG	GGCAGTCTG	3540
CTCAGTCCTA	ATTAAACGGG	ATTCCGGTGA	ACTACTTCCG	CTAGCCGTCC	GTATGGGGGC	3600
TATTGCTCTC	ATGAGGATAC	AGGGTCCGGT	TGTCCATGCC	CAATCAGGGA	TGTTACTGAC	3660
AGGGGCCAAT	GCAAGGGGGA	TGGATCTTGG	CACATATACCA	GGAGACTGCG	GGGCACCATA	3720
CGTCCACAAG	CGCGGCAATG	ACTGGGTTGT	GTGTGGAGTC	CACGCTGCAG	CCACAAAGTC	3780
AGGCCAACACC	GTGGTCTGCG	CTGTACAGGC	TGGAGAGGGC	GAAACCGCAC	TAGAAGGTGG	3840
AGACAAAGGG	CATTATGCCG	GCCACGAGAT	TGTGAGGTAT	GGAAGTGGCC	CAGCACTGTC	3900
AACTAAAACA	AAATTCTGGA	GGTCTCTCCC	AGAACCCTG	CCCCCGGAG	TATATGAGCC	3960
AGCATACCTG	GGGGGCAAGG	ACCCCGGTGT	ACAGAATGGC	CCATCCCTAC	AACAGGTACT	4020
ACGTGACCAA	CTGAAACCTT	TTGCGGACCC	CCGCGGCCGC	ATGCCTGAGC	CTGGCCTACT	4080
GGAGGCTGCG	GTTGAGACTG	TAACATCCAT	GTTAGAACAG	ACAATGGATA	CCCCAAGCCC	4140
GTGGTCTTAC	GCTGATGCCT	GCCATCTCT	TGACAAAAT	ACTAGTTCGG	GGTACCCTCA	4200
CCATAAAAGG	AAGATGATG	ATTGGAATGG	CACCACCTTC	GTTGGAGAGC	TCGGTGAGCA	4260
AGCTGCACAC	GCCAAACAATA	TGTATGAGAA	TGCTAAACAT	ATGAAACCCA	TTTACACTGC	4320
AGCCTTAAAA	GATGAAGTAG	TCAAGCCAGA	AAAGATTAT	CAAAAAGTCA	AGAAGCGTCT	4380
ACTATGGGGC	GCCGATCTCG	GAACAGTGGT	CAGGGCCGCC	CGGGCTTTTG	GCCCATTTTG	4440
TGACGCTATA	AAATCACATG	TCATCAAATT	GCCAAATAAA	GTTGGCATGA	ACACAATAGA	4500
AGATGGCCCC	CTCATCTATG	CTGAGCATGC	TAAATATAAG	AATCATTTTG	ATGCAGATTA	4560
TACAGCATGG	GACTCBAAC	AAAATAGACA	AATTATGACA	GAATCTTCT	CCATTATGTC	4620
GCGCCTTACG	GCCTCACCAG	AATTGGCCGA	GGTTGTGGCC	CAAGATTGTC	TAGCACCATC	4680
TGAGATGGAT	GTAGGTGATT	ATGTCATCAG	GGTCAAAGAG	GGGCTGCCAT	CTGGATTCCC	4740
ATGTACTTCC	CAGGTGAACA	GCATAAATCA	CTGGATAATT	ACTCTCTGTG	CACTGTCTGA	4800
GGCCACTGGT	TTATCACCTG	ATGTGGTGCA	ATCCATGTCA	TATTTCTCAT	TTTATGGTGA	4860
GATATGAGAT	GTGTCAACTG	ACATGATATT	TGACCCAGCC	CGCCTCACTC	AAATTCTCAA	4920
GGAATATGGC	CTCAAACCAA	CAAGGCCTGA	CAAAACAGAA	GGACCAATAC	AAGTGAGGAA	4980
AAATGTGGAT	GGACTGGTCT	TCTTGCGGCG	CACCATTTC	CGTGATGCGG	CAGGGTTCCA	5040
AGGCAGGTGA	GATAGGGCTT	CGATTGAACG	CCAAATCTTC	TGGACCCGCG	GGCCCAATCA	5100
TTCAGATCCA	TCAGAGACTC	TAGTGCCACA	CACCTAAAGA	AAAATACAGT	TGATTTCACT	5160
TCTAGGGGAA	GCTTCACTCC	ATGGTGAGAA	ATTTTACAGA	AAGATTTCCT	GCAAGGTCAT	5220
ACATGAAATC	AAGACTGGTG	GATTGGAAAT	GTATGTCCCA	GGATGGCAGG	CCATGTTCCG	5280
CTGGATGCGC	TTCCATGACC	TCGGATTGTG	GACAGGAGAT	CGCGATCTTC	TGCCCGAATT	5340
CGTAAATGAT	GATGCGTCT	AAGGACGCTA	CATCAAGCGT	GGATGGCGCT	AGTGGCGCTG	5400
GTCAGTTGGT	ACCGAGGTT	AATGCTTCTG	ACCTCTTTC	AATGGATCCT	GTCAGAGGTT	5460
CTTCGACAGC	AGTCGCGACT	GCTGGACAAG	TTAATCCTAT	TGATCCCTGG	ATAATTAATA	5520
ATTTTGTGCA	AGCCCCCAA	GGTGAATTTA	CTATTTCCCC	AAATAATACC	CCCGGTGATG	5580
TTTTGTTTGA	TTTGAGTTTG	GGTCCCCATC	TTAATCCTTT	CTTGCTCCAT	CTATCACAAA	5640
TGTATAATGG	TTGGGTTGGT	AACATGAGAG	TCAGGATTAT	GCTAGCTGGT	AATGCCTTTA	5700
CTGCGGGGAA	GATAATAGTT	TCTGTCATAC	CCCCTGGTTT	TGGTTCACAT	AATCTTACTA	5760
TAGCACAAGC	AACCTCTCTT	CCACATGTGA	TGCTGATGT	TAGGACTCTA	GACCCCATTG	5820
AGGTGCCCTT	GGAAGATGTT	AGGAATGTTC	TCTTTTATAA	TAATGATAGA	AATCAACAAA	5880
CCATGCGCCT	TGTGTGCATG	CTGTACACCC	CCCTCCGCAC	TGGTGGTGGT	ACTGGTGATT	5940
CTTTTGTAGT	TGCGGGGCGA	GTTATGACTT	GCCCCAGTCC	TGATTTTAAAT	TTCTTGTTTT	6000
TAGTCCCTCC	TACGGTGGAG	CAGAAAACCA	GGCCCTTCAC	ACTCCCAAT	CTGCCATTGA	6060
GTTCTCTGTC	TAACTCACGT	GCCCCCTCTC	CAATCAGTAG	TATGGGCATT	TCCCCAGACA	6120
ATGTCAGAG	TGTGCAGTTC	CAAAATGGTC	GGTGTACTCT	GGATGGCCGC	CTGGTTGGCA	6180
CCACCCAGT	TTCAATTGTCA	CATGTTGCCA	AGATAAGAGG	GACCTCCAAT	GGCACTGTAA	6240
TCAACCTTAC	TGAATTGGAT	GGCACACCTT	TTACCCCTTT	TGAGGGCCCT	GCCCCATTG	6300
GGTTTCCAGA	CCTCGGTGGT	TGTGATTGGC	ATATCAATAT	GACACAGTTT	GGCCATTCTA	6360
GCCAGACCCA	GTATGATGTA	GACACCACCC	CTGACACTTT	TGTCCCCCAT	CTTGGTTCAA	6420
TTCAGGCAAA	TGGCATTGGC	AGTGGTAATT	ATGTTGGTGT	TCTTAGCTGG	ATTCCCCCCC	6480

CATCACACCC	GTCTGGCTCC	CAAGTTGACC	TTTGAAGAT	CCCCAATTAT	GGGTCAAGTA	6540
TTACGGAGGC	AACACATCTA	GCCCCTTCTG	TATACCCCCC	TGGTTTCGGA	GAGGTATTGG	6600
TCTTTTTCAT	GTCAAAAATG	CCAGGTCCTG	GTGCTTATAA	TTTGCCCTGT	CTATTACCAC	6660
AAGAGTACAT	TTCACATCTT	GCTAGTGAAC	AAGCCCCTAC	TGTAGGTGAG	GCTGCCCTGC	6720
TCCACTATGT	TGACCCTGAT	ACCGGTCGGA	ATCTTGGGGA	ATTCAAAGCA	TACCCTGATG	6780
GTTTCCTCAC	TTGTGTCCCC	AATGGGGCTA	GCTCGGGTCC	ACAACAGCTG	CCGATCAATG	6840
GGGTCTTTGT	CTTTGTTTCA	TGGGTGTCCA	GATTTTATCA	ATTAAAGCCT	GTGGGAACTG	6900
CCAGCTCGGC	AAGAGGTAGG	CTTGGTCTGC	GCCGATAATG	GCCCAAGCCA	TAATTGGTGC	6960
AATTGCTGCT	TCCACAGCAG	GTAGTGCTCT	GGGAGCGGCG	ATACAGGTTG	GTGGCGAAGC	7020
GGCCCTCCAA	AGCCAAAGGT	ATCAACAAAA	TTTGCAACTG	CAAGAAAATT	CTTTTAAACA	7080
TGACAGGGAA	ATGATTGGGT	ATCAGGTTGA	AGCTTCAAAT	CAATTATTGG	CTAAAAATTT	7140
GGCAACTAGA	TATCACTCC	TCCGTGCTGG	GGGTTTGACC	AGTGCTGATG	CAGCAAGATC	7200
TGTGGCAGGA	GCTCCAGTCA	CCCGCATTGT	AGATTGGAAT	GGCGTGAGAG	TGTCTGCTCC	7260
CGAGTCTCT	GCTACCACAT	TGAGATCCGG	TGGCTTCATG	TGAGTTCCCA	TACCATTTCG	7320
CTCTAAGCAA	AAACAGGTTT	AATCATCTGG	TATTAGTAAT	CCAAATATT	CCCCTTCATC	7380
CATTTCTCGA	ACCACTAGTT	GGGTCGAGTC	ACAAAACCTCA	TCGAGATTG	GAAATCTTTC	7440
TCCATACCAC	GCGGAGGCTC	TCAATACAGT	GTGGTTGACT	CCACCCGTT	CAACAGCCTC	7500
TTCTACACTG	TCTTCTGTGC	CACGTGGTTA	TTTCAATACA	GACAGGTTGC	CATTATTTCG	7560
AAATAATAGG	CGATGATGTT	GTAATATGAA	ATGTGGGCAT	CATATTCAAT	TAATTAGGTT	7620
TAATTAGGTT	TAATTGATG	TTAAAAAAA	AAAAAAA	AAAAAAA	AAAAAAA	7680
AAAAAAA	AAAAAAA	AAAAAAA	AAAAAAA	AAAAAAA	AAAAAAA	7740
AAAAAAA	AAA					7753

(2) INFORMATION FOR SEQ ID NO:2:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1738 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met	Ala	Leu	Gly	Leu	Ile	Gly	Gln	Val	Pro	Ala	Pro	Lys	Ala	Thr	Ser
1				5					10					15	
Val	Asp	Val	Pro	Lys	Gln	Gln	Arg	Asp	Arg	Pro	Pro	Arg	Thr	Val	Ala
		20						25					30		
Glu	Val	Gln	Gln	Asn	Leu	Arg	Trp	Thr	Glu	Arg	Pro	Gln	Asp	Gln	Asn
		35					40					45			
Val	Lys	Thr	Trp	Asp	Glu	Leu	Asp	His	Thr	Thr	Lys	Gln	Gln	Ile	Leu
	50					55					60				
Asp	Glu	His	Ala	Glu	Trp	Phe	Asp	Ala	Gly	Gly	Leu	Gly	Pro	Ser	Thr
	65				70					75				80	
Leu	Pro	Thr	Ser	His	Glu	Arg	Tyr	Thr	His	Glu	Asn	Asp	Glu	Gly	His
			85						90					95	
Gln	Val	Lys	Trp	Ser	Ala	Arg	Glu	Gly	Val	Asp	Leu	Gly	Ile	Ser	Gly
		100						105					110		
Leu	Thr	Thr	Val	Ser	Gly	Pro	Glu	Trp	Asn	Met	Cys	Pro	Leu	Pro	Pro
	115						120					125			
Val	Asp	Gln	Arg	Ser	Thr	Thr	Pro	Ala	Thr	Glu	Pro	Thr	Ile	Gly	Asp
	130						135					140			
Met	Ile	Glu	Phe	Tyr	Glu	Gly	His	Ile	Tyr	His	Tyr	Ala	Ile	Tyr	Ile
145					150				155						160

Gly Gln Gly Lys Thr Val Gly Val His Ser Pro Gln Ala Ala Phe S r
 165 170 175
 Ile Thr Arg Ile Thr Ile Gln Pro Ile Ser Ala Trp Trp Arg Val Cys
 180 185 190
 Tyr Val Pro Gln Pro Lys Gln Arg Leu Thr Tyr Asp Gln Leu Lys Glu
 195 200 205
 Leu Glu Asn Glu Pro Trp Pro Tyr Ala Ala Val Thr Asn Asn Cys Phe
 210 215 220
 Glu Phe Cys Cys Gln Val Met Cys Leu Glu Asp Thr Trp Leu Gln Arg
 225 230 235 240
 Lys Leu Ile Ser Ser Gly Arg Phe Tyr His Pro Thr Gln Asp Trp Ser
 245 250 255
 Arg Asp Thr Pro Glu Phe Gln Gln Asp Ser Lys Leu Glu Met Val Arg
 260 265 270
 Asp Ala Val Leu Ala Ala Ile Asn Gly Leu Val Ser Arg Pro Phe Lys
 275 280 285
 Asp Leu Leu Gly Lys Leu Lys Pro Leu Asn Val Leu Asn Leu Leu Ser
 290 295 300
 Asn Cys Asp Trp Thr Phe Met Gly Val Val Glu Met Val Val Leu Leu
 305 310 315 320
 Leu Glu Leu Phe Gly Ile Phe Trp Asn Pro Pro Asp Val Ser Asn Phe
 325 330 335
 Ile Ala Ser Leu Leu Pro Asp Phe His Leu Gln Gly Pro Glu Asp Leu
 340 345 350
 Ala Arg Asp Leu Val Pro Ile Val Leu Gly Gly Ile Gly Leu Ala Ile
 355 360 365
 Gly Phe Thr Arg Asp Lys Val Ser Lys Met Met Lys Asn Ala Val Asp
 370 375 380
 Gly Leu Arg Ala Ala Thr Gln Leu Gly Gln Tyr Gly Leu Glu Ile Phe
 385 390 395 400
 Ser Leu Leu Lys Lys Tyr Phe Phe Gly Gly Asp Gln Thr Glu Lys Thr
 405 410 415
 Leu Lys Asp Ile Glu Ser Ala Val Ile Asp Met Glu Val Leu Ser Ser
 420 425 430
 Thr Ser Val Thr Gln Leu Val Arg Asp Lys Gln Ser Ala Arg Ala Tyr
 435 440 445
 Met Ala Ile Leu Asp Asn Glu Glu Glu Lys Ala Arg Lys Leu Ser Val
 450 455 460
 Arg Asn Ala Asp Pro His Val Val Ser Ser Thr Asn Ala Leu Ile Ser
 465 470 475 480
 Arg Ile Ser Met Ala Arg Ala Ala Leu Ala Lys Ala Gln Ala Glu Met
 485 490 495
 Thr Ser Arg Met Arg Pro Val Val Ile Met Met Cys Gly Pro Pro Gly
 500 505 510

Ile Gly Lys Thr Lys Ala Ala Glu His Leu Ala Lys Arg Leu Ala Asn
 515 520 525
 Glu Ile Arg Pro Gly Gly Lys Val Gly Leu Val Pro Arg Glu Ala Val
 530 535 540
 Asp His Trp Asp Gly Tyr His Gly Glu Glu Val Met Leu Trp Asp Asp
 545 550 555 560
 Tyr Gly Met Thr Lys Ile Gln Glu Asp Cys Asn Lys Leu Gln Ala Ile
 565 570 575
 Ala Asp Ser Ala Pro Leu Thr Leu Asn Cys Asp Arg Ile Glu Asn Lys
 580 585 590
 Gly Met Gln Phe Val Ser Asp Ala Ile Val Ile Thr Thr Asn Ala Pro
 595 600 605
 Gly Pro Ala Pro Val Asp Phe Val Asn Leu Gly Pro Val Cys Arg Arg
 610 615 620
 Val Asp Phe Leu Val Tyr Cys Thr Ala Pro Glu Val Glu His Thr Arg
 625 630 635 640
 Lys Val Ser Pro Gly Asp Thr Thr Ala Leu Lys Asp Cys Phe Lys Pro
 645 650 655
 Asp Phe Ser His Leu Lys Met Glu Leu Ala Pro Gln Gly Gly Phe Asp
 660 665 670
 Asn Gln Gly Asn Thr Pro Phe Gly Lys Gly Val Met Lys Pro Thr Thr
 675 680 685
 Ile Asn Arg Leu Leu Ile Gln Ala Val Ala Leu Thr Met Glu Arg Gln
 690 695 700
 Asp Glu Phe Gln Leu Gln Gly Pro Thr Tyr Asp Phe Asp Thr Asp Arg
 705 710 715 720
 Val Ala Ala Phe Thr Arg Met Ala Arg Ala Asn Gly Leu Gly Leu Ile
 725 730 735
 Ser Met Ala Ser Leu Gly Lys Lys Leu Arg Ser Val Thr Thr Ile Glu
 740 745 750
 Gly Leu Lys Asn Ala Leu Ser Gly Tyr Lys Ile Ser Lys Cys Ser Ile
 755 760 765
 Gln Trp Gln Ser Arg Val Tyr Ile Ile Glu Ser Asp Gly Ala Ser Val
 770 775 780
 Gln Ile Lys Glu Asp Lys Gln Ala Leu Thr Pro Leu Gln Gln Thr Ile
 785 790 795 800
 Asn Thr Ala Ser Leu Ala Ile Thr Arg Leu Lys Ala Ala Arg Ala Val
 805 810 815
 Ala Tyr Ala Ser Cys Phe Gln Ser Ala Ile Thr Thr Ile Leu Gln Met
 820 825 830
 Ala Gly Ser Ala Leu Val Ile Asn Arg Ala Val Lys Arg Met Phe Gly
 835 840 845
 Thr Arg Thr Ala Ala Met Ala Leu Glu Gly Pro Gly Lys Glu His Asn
 850 855 860

Cys Arg Val His Lys Ala Lys Glu Ala Gly Lys Gly Pro Ile Gly His
 865 870 875 880
 Asp Asp Met Val Glu Arg Phe Gly Leu Cys Glu Thr Glu Glu Glu Glu
 885 890 895
 Ser Glu Asp Gln Ile Gln Met Val Pro Ser Asp Ala Val Pro Glu Gly
 900 905 910
 Lys Asn Lys Gly Lys Thr Lys Lys Gly Arg Gly Arg Lys Asn Asn Tyr
 915 920 925
 Asn Ala Phe Ser Arg Arg Gly Leu Ser Asp Glu Glu Tyr Glu Glu Tyr
 930 935 940
 Lys Lys Ile Arg Glu Glu Lys Asn Gly Asn Tyr Ser Ile Gln Glu Tyr
 945 950 955 960
 Leu Glu Asp Arg Gln Arg Tyr Glu Glu Glu Leu Ala Glu Val Gln Ala
 965 970 975
 Gly Gly Asp Gly Gly Ile Gly Glu Thr Glu Met Glu Ile Arg His Arg
 980 985 990
 Val Phe Tyr Lys Ser Lys Ser Lys Lys His Gln Gln Glu Gln Arg Arg
 995 1000 1005
 Gln Leu Gly Leu Val Thr Gly Ser Asp Ile Arg Lys Arg Lys Pro Ile
 1010 1015 1020
 Asp Trp Thr Pro Pro Lys Asn Glu Trp Ala Asp Asp Asp Arg Glu Val
 1025 1030 1035 1040
 Asp Tyr Asn Glu Lys Ile Asn Phe Glu Ala Pro Pro Thr Leu Trp Ser
 1045 1050 1055
 Arg Val Thr Lys Phe Gly Ser Gly Trp Gly Phe Trp Val Ser Pro Thr
 1060 1065 1070
 Val Phe Ile Thr Thr Thr His Val Val Pro Thr Gly Val Lys Glu Phe
 1075 1080 1085
 Phe Gly Glu Pro Leu Ser Ser Ile Ala Ile His Gln Ala Gly Glu Phe
 1090 1095 1100
 Thr Gln Phe Arg Phe Ser Lys Lys Met Arg Pro Asp Leu Thr Gly Met
 1105 1110 1115 1120
 Val Leu Glu Glu Gly Cys Pro Glu Gly Thr Val Cys Ser Val Leu Ile
 1125 1130 1135
 Lys Arg Asp Ser Gly Glu Leu Leu Pro Leu Ala Val Arg Met Gly Ala
 1140 1145 1150
 Ile Ala Ser Met Arg Ile Gln Gly Arg Leu Val His Gly Gln Ser Gly
 1155 1160 1165
 Met Leu Leu Thr Gly Ala Asn Ala Lys Gly Met Asp Leu Gly Thr Ile
 1170 1175 1180
 Pro Gly Asp Cys Gly Ala Pro Tyr Val His Lys Arg Gly Asn Asp Trp
 1185 1190 1195 1200
 Val Val Cys Gly Val His Ala Ala Ala Thr Lys Ser Gly Asn Thr Val
 1205 1210 1215

Val Cys Ala Val Gln Ala Gly Glu Gly Glu Thr Ala Leu Glu Gly Gly
 1220 1225 1230
 Asp Lys Gly His Tyr Ala Gly His Glu Ile Val Arg Tyr Gly Ser Gly
 1235 1240 1245
 Pro Ala Leu Ser Thr Lys Thr Lys Phe Trp Arg Ser Ser Pro Glu Pro
 1250 1255 1260
 Leu Pro Pro Gly Val Tyr Glu Pro Ala Tyr Leu Gly Gly Lys Asp Pro
 1265 1270 1275 1280
 Arg Val Gln Asn Gly Pro Ser Leu Gln Gln Val Leu Arg Asp Gln Leu
 1285 1290 1295
 Lys Pro Phe Ala Asp Pro Arg Gly Arg Met Pro Glu Pro Gly Leu Leu
 1300 1305 1310
 Glu Ala Ala Val Glu Thr Val Thr Ser Met Leu Glu Gln Thr Met Asp
 1315 1320 1325
 Thr Pro Ser Pro Trp Ser Tyr Ala Asp Ala Cys Gln Ser Leu Asp Lys
 1330 1335 1340
 Thr Thr Ser Ser Gly Tyr Pro His His Lys Arg Lys Asn Asp Asp Trp
 1345 1350 1355 1360
 Asn Gly Thr Thr Phe Val Gly Glu Leu Gly Glu Gln Ala Ala His Ala
 1365 1370 1375
 Asn Asn Met Tyr Glu Asn Ala Lys His Met Lys Pro Ile Tyr Thr Ala
 1380 1385 1390
 Ala Leu Lys Asp Glu Leu Val Lys Pro Glu Lys Ile Tyr Gln Lys Val
 1395 1400 1405
 Lys Lys Arg Leu Leu Trp Gly Ala Asp Leu Gly Thr Val Val Arg Ala
 1410 1415 1420
 Ala Arg Ala Phe Gly Pro Phe Cys Asp Ala Ile Lys Ser His Val Ile
 1425 1430 1435 1440
 Lys Leu Pro Ile Lys Val Gly Met Asn Thr Ile Glu Asp Gly Pro Leu
 1445 1450 1455
 Ile Tyr Ala Glu His Ala Lys Tyr Lys Asn His Phe Asp Ala Asp Tyr
 1460 1465 1470
 Thr Ala Trp Asp Ser Thr Gln Asn Arg Gln Ile Met Thr Glu Ser Phe
 1475 1480 1485
 Ser Ile Met Ser Arg Leu Thr Ala Ser Pro Glu Leu Ala Glu Val Val
 1490 1495 1500
 Ala Gln Asp Leu Leu Ala Pro Ser Glu Met Asp Val Gly Asp Tyr Val
 1505 1510 1515 1520
 Ile Arg Val Lys Glu Gly Leu Pro Ser Gly Phe Pro Cys Thr Ser Gln
 1525 1530 1535
 Val Asn Ser Ile Asn His Trp Ile Ile Thr Leu Cys Ala Leu Ser Glu
 1540 1545 1550
 Ala Thr Gly Leu Ser Pro Asp Val Val Gln Ser Met Ser Tyr Phe Ser
 1555 1560 1565

Phe Tyr Gly Asp Asp Glu Ile Val Ser Thr Asp Ile Asp Phe Asp Pro
 1570 1575 1580
 Ala Arg Leu Thr Gln Ile Leu Lys Glu Tyr Gly Leu Lys Pro Thr Arg
 1585 1590 1595 1600
 Pro Asp Lys Thr Glu Gly Pro Ile Gln Val Arg Lys Asn Val Asp Gly
 1605 1610 1615
 Leu Val Phe Leu Arg Arg Thr Ile Ser Arg Asp Ala Ala Gly Phe Gln
 1620 1625 1630
 Gly Arg Leu Asp Arg Ala Ser Ile Glu Arg Gln Ile Phe Trp Thr Arg
 1635 1640 1645
 Gly Pro Asn His Ser Asp Pro Ser Glu Thr Leu Val Pro His Thr Gln
 1650 1655 1660
 Arg Lys Ile Gln Leu Ile Ser Leu Leu Gly Glu Ala Ser Leu His Gly
 1665 1670 1675 1680
 Glu Lys Phe Tyr Arg Lys Ile Ser Ser Lys Val Ile His Glu Ile Lys
 1685 1690 1695
 Thr Gly Gly Leu Glu Met Tyr Val Pro Gly Trp Gln Ala Met Phe Arg
 1700 1705 1710
 Trp Met Arg Phe His Asp Leu Gly Leu Trp Thr Gly Asp Arg Asp Leu
 1715 1720 1725
 Leu Pro Glu Phe Val Asn Asp Asp Gly Val
 1730 1735

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 530 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

Met Met Met Ala Ser Lys Asp Ala Thr Ser Ser Val Asp Gly
 1 5 10
 Ala Ser Gly Ala Gly Gln Leu Val Pro Glu Val Asn Ala Ser Asp Pro
 15 20 25 30
 Leu Ala Met Asp Pro Val Ala Gly Ser Ser Thr Ala Val Ala Thr Ala
 35 40 45
 Gly Gln Val Asn Pro Ile Asp Pro Trp Ile Ile Asn Asn Phe Val Gln
 50 55 60
 Ala Pro Gln Gly Glu Phe Thr Ile Ser Pro Asn Asn Thr Pro Gly Asp
 65 70 75
 Val Leu Phe Asp Leu Ser Leu Gly Pro His Leu Asn Pro Phe Leu Leu
 80 85 90
 His Leu Ser Gln Met Tyr Asn Gly Trp Val Gly Asn Met Arg Val Arg
 95 100 105 110

Ile Met Leu Ala Gly Asn Ala Phe Thr Ala Gly Lys Ile Ile Val Ser
 115 120 125
 Cys Ile Pro Pro Gly Phe Gly Ser His Asn Leu Thr Ile Ala Gln Ala
 130 135 140
 Thr Leu Phe Pro His Val Ile Ala Asp Val Arg Thr Leu Asp Pro Ile
 145 150 155
 Glu Val Pro Leu Glu Asp Val Arg Asn Val Leu Phe His Asn Asn Asp
 160 165 170
 Arg Asn Gln Gln Thr Met Arg Leu Val Cys Met Leu Tyr Thr Pro Leu
 175 180 185 190
 Arg Thr Gly Gly Gly Thr Gly Asp Ser Phe Val Val Ala Gly Arg Val
 195 200 205
 Met Thr Cys Pro Ser Pro Asp Phe Asn Phe Leu Phe Leu Val Pro Pro
 210 215 220
 Thr Val Glu Gln Lys Thr Arg Pro Phe Thr Leu Pro Asn Leu Pro Leu
 225 230 235
 Ser Ser Leu Ser Asn Ser Arg Ala Pro Leu Pro Ile Ser Ser Met Gly
 240 245 250
 Ile Ser Pro Asp Asn Val Gln Ser Val Gln Phe Gln Asn Gly Arg Cys
 255 260 265 270
 Thr Leu Asp Gly Arg Leu Val Gly Thr Thr Pro Val Ser Leu Ser His
 275 280 285
 Val Ala Lys Ile Arg Gly Thr Ser Asn Gly Thr Val Ile Asn Leu Thr
 290 295 300
 Glu Leu Asp Gly Thr Pro Phe His Pro Phe Glu Gly Pro Ala Pro Ile
 305 310 315
 Gly Phe Pro Asp Leu Gly Gly Cys Asp Trp His Ile Asn Met Thr Gln
 320 325 330
 Phe Gly His Ser Ser Gln Thr Gln Tyr Asp Val Asp Thr Thr Pro Asp
 335 340 345 350
 Thr Phe Val Pro His Leu Gly Ser Ile Gln Ala Asn Gly Ile Gly Ser
 355 360 365
 Gly Asn Tyr Val Gly Val Leu Ser Trp Ile Ser Pro Pro Ser His Pro
 370 375 380
 Ser Gly Ser Gln Val Asp Leu Trp Lys Ile Pro Asn Tyr Gly Ser Ser
 385 390 395
 Ile Thr Glu Ala Thr His Leu Ala Pro Ser Val Tyr Pro Pro Gly Phe
 400 405 410
 Gly Glu Val Leu Val Phe Phe Met Ser Lys Met Pro Gly Pro Gly Ala
 415 420 425 430
 Tyr Asn Leu Pro Cys Leu Leu Pro Gln Glu Tyr Ile Ser His Leu Ala
 435 440 445
 Ser Glu Gln Ala Pro Thr Val Gly Glu Ala Ala Leu Leu His Tyr Val
 450 455 460

Asp Pro Asp Thr Gly Arg Asn Leu Gly Glu Phe Lys Ala Tyr Pro Asp
 465 470 475
 Gly Phe Leu Thr Cys Val Pro Asn Gly Ala Ser Ser Gly Pro Gln Gln
 480 485 490
 Leu Pro Ile Asn Gly Val Phe Val Phe Val Ser Trp Val Ser Arg Phe
 495 500 505 510
 Tyr Gln Leu Lys Pro Val Gly Thr Ala Ser Ser Ala Arg Gly Arg Leu
 515 520 525
 Gly Leu Arg Arg
 530

(2) INFORMATION FOR SEQ ID NO:4:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 212 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

 Met Ala Gln Ala Ile Ile
 1 5
 Gly Ala Ile Ala Ala Ser Thr Ala Gly Ser Ala Leu Gly Ala Gly Ile
 10 15 20
 Gln Val Gly Gly Glu Ala Ala Leu Gln Ser Gln Arg Tyr Gln Gln Asn
 25 30 35
 Leu Gln Leu Gln Glu Asn Ser Phe Lys His Asp Arg Glu Met Ile Gly
 40 45 50
 Tyr Gln Val Glu Ala Ser Asn Gln Leu Leu Ala Lys Asn Leu Ala Thr
 55 60 65 70
 Arg Tyr Ser Leu Leu Arg Ala Gly Gly Leu Thr Ser Ala Asp Ala Ala
 75 80 85
 Arg Ser Val Ala Gly Ala Pro Val Thr Arg Ile Val Asp Trp Asn Gly
 90 95 100
 Val Arg Val Ser Ala Pro Glu Ser Ser Ala Thr Thr Leu Arg Ser Gly
 105 110 115
 Gly Phe Met Ser Val Pro Ile Pro Phe Ala Ser Lys Gln Lys Gln Val
 120 125 130
 Gln Ser Ser Gly Ile Ser Asn Pro Asn Tyr Ser Pro Ser Ser Ile Ser
 135 140 145 150
 Arg Thr Thr Ser Trp Val Glu Ser Gln Asn Ser Ser Arg Phe Gly Asn
 155 160 165
 Leu Ser Pro Tyr His Ala Glu Ala Leu Asn Thr Val Trp Leu Thr Pro
 170 175 180
 Pro Gly Ser Thr Ala Ser Ser Thr Leu Ser Ser Val Pro Arg Gly Tyr
 185 190 195

Phe Asn Thr Asp Arg Leu Pro Leu Phe Ala Asn Asn Arg Arg
 200 205 210

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 551 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human calicivirus Sapporo

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..549

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

```

TGTGATGCTG CCACCACGCT TATAGCCACC GCGGCTTTTA AGGCCGTGGC TACNAGGCTA      60
CAGGTGGTGA CACCAATGAC ACCAGTTGCT GTTGGCATTAC ATATGGACTC TGTTGAGATG      120
CAAGTGATGA ATGACTCTTT AAAGGGGGGT GTTCTTTTACT GTTTGGATTA TTCCAAATGG      180
GATTCCACAC AAAACCCTGC AGTGACAGCA GCCTCCCTGG CAATATTGGA GAGATTGCT      240
GAGCCCCATC CAATTGTGTC TTGTGCCATT GAGGCTCTTT CCTCCCCTGC AGAGGGCTAT      300
GTCAATGATA TCAAATTTGT GACACGCGGC GGCCTACCAT CTGGGATGCC ATTTACATCT      360
GTCGTCAATT CTATCAACCA TATGATATAC GTGGCGGCAG CCATCCTGCA GGCATACGAA      420
AGCCACAATG TCCCATATAC TGGAAACGTC TTCCAAGTGG AGACCGTTCA CACGTATGGT      480
GATGATTGCA TGTACAGCGT GTGCCCTGCC ACTGCATCAA TTTCCACAC TGTGCTTGCA      540
AACCTAACGT C                                                                551

```

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 183 amino acids
- (B) TYPE: amino acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

```

Cys Asp Ala Ala Thr Thr Leu Ile Ala Thr Ala Ala Phe Lys Ala Ala
 1           5           10           15
Val Xaa Arg Leu Gln Val Val Thr Pro Met Thr Pro Val Ala Val Gly
          20           25           30
Ile Asn Met Asp Ser Val Gln Met Gln Val Met Asn Asp Ser Leu Lys
          35           40           45

```

Gly Gly Val Leu Tyr Cys Leu Asp Tyr Ser Lys Trp Asp Ser Thr Gln
 50 55 60
 Asn Pro Ala Val Thr Ala Ala Ser Leu Ala Ile Leu Glu Arg Phe Ala
 65 70 75 80
 Glu Pro His Pro Ile Val Ser Cys Ala Ile Glu Ala Leu Ser Ser Pro
 85 90 95
 Ala Glu Gly Tyr Val Asn Asp Ile Lys Phe Val Thr Arg Gly Gly Leu
 100 105 110
 Pro Ser Gly Met Pro Phe Thr Ser Val Val Asn Ser Ile Asn His Met
 115 120 125
 Ile Tyr Val Ala Ala Ala Ile Leu Gly Ala Tyr Glu Ser His Asn Val
 130 135 140
 Pro Tyr Thr Gly Asn Val Phe Gln Val Glu Thr Val His Thr Tyr Gly
 145 150 155 160
 Asp Asp Cys Met Tyr Ser Val Cys Pro Ala Thr Ala Ser Ile Phe His
 165 170 175
 Thr Val Leu Ala Asn Leu Thr
 180

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 148 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

TGTGATGCTG CCACCACGCT TATAGCCACC GCGGCTTTTA AGGCCGTGGC TACAGGCTAC	60
AGGTGGTGAC ACCAATGACA CCAGTTGCTG TTGGCATTAA CATGGACTCT GTTCAGATGC	120
AAGTGATGAA TGACTCTTTA AAGGGGGG	148

(2) INFORMATION FOR SEQ ID NO:8

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 449 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ATGGACTCTG TTCAGATGCA AGTGATGAAT GACTCTTTAA AGGGGGGTGT TCTTTACTGT	60
TTGGATTATT CCAAATGGGA TTCCACACAA AACCTGCGAG TGACAGCAGC CTCCCTGGCA	120
ATATTGGAGA GATTTGCTGA GCCCCATCCA ATTGTGTCTT GTGCCATTGA GGCTCTTTCC	180
TCCCCTGCAG AGGGCTATGT CAATGATATC AAATTGTGA CACGCGCGG CCTACCATCT	240

GGGATGCCAT TTACATCTGT CGTCAATTCT ATCAACCATA TGATATACGT GCGGGCAGCC	300
ATCCTGCAGG CATACGAAAG CCACAATGTC CCATATACTG GAAACGTCTT CCAAGTGGAG	360
ACCGTTCACA CGTATGGTGA TGATTGCATG TACAGCGTGT GCCCTGCCAC TGCATCAATT	420
TTCCACACTG TGCTTGCAA CCTAACGTC	449

(2) INFORMATION FOR SEQ ID NO:9:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 446 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human calicivirus Saporro (Day care)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

ATGGACTCTG TTCAGATGCA AGTGATGAAT GACTCTTTAA AGGGAGGTGT TCTCTACTGC	60
CTGGATTACT CCAAATGGGA CTCCACACAA AATGCTGCAG TGACAGCAGC ATCCCTNNCA	120
ATATTGGAGA GATTTGCTGA ACCCCACCCA ATTGTGTCTT GTGCCATTGA GGCCCTGNNC	180
TCNNCTGCAG AGGGTTACGT TAATGATATC AAGTTTGTA CACGTGGCGG CCTACCATGT	240
GGGATGCCAT TCACATCTGT TGTCAATTCC ATCAACCACA TNATATACGT GGCAGCCGCC	300
ATCCTGCAGG CATACGAAAG CCACAATGTT CCATACACTG GAAATGTCTT CCAAGTGGAG	360
ACTGTTTACA CGTATGGTGA CGATTGCATG TACAGCGTGT GCCCTGCCAC CGCATCAATT	420
TTCCACACTG TACTTGCAA CCTAAC	446

(2) INFORMATION FOR SEQ ID NO:10:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 434 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: human calicivirus Houston

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 3..434

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

GGCCATGTTA TAGTGGTGTT CACATGAAAG ATGGCGACAA GATGTTGATA GATGCCAATC	60
TTCCTTACAA CCAGAAATTA ACTACTATGA TTCATGAGAC TAGGCATAGG ATAGGACAGT	120
ATATAGATAA TACTTTTGA AAGACATTTA GACATGGATT GACAAAACCT GCTGACAAGA	180

CTGTAGATTT GATCTATAAG ACATTGAATT ATGATGATTT TCTGGCAATA ATGCTAATCA 240
 TATATGGGCA AAAGTCGGCC ACTAATACGG AGTTGCAATT CTTGATGGAG AAACCTTAGAG 300
 GTTATGAATC TACAATGGAT GACATAGGGA AAGTCTATGG AGATGATAAA ATGAGAGATA 360
 TAATCAAGAA TATTTCTGAT GATGACATAA AGAGTCTTTT AGGGGAGATA AATAGTGATT 420
 ATTCTGGTAA GNAT 434

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 144 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Pro Cys Tyr Ser Gly Val His Met Lys Asp Gly Asp Lys Met Leu Ile
 1 5 10 15
 Asp Ala Asn Leu Pro Tyr Asn Gln Lys Leu Thr Thr Met Ile His Glu
 20 25 30
 Thr Arg His Arg Ile Gly Gln Tyr Ile Asp Asn Thr Phe Gly Lys Thr
 35 40 45
 Phe Arg His Gly Leu Thr Lys Pro Ala Asp Lys Thr Val Asp Leu Ile
 50 55 60
 Tyr Lys Thr Leu Asn Tyr Asp Asp Phe Leu Ala Ile Met Leu Ile Ile
 65 70 75 80
 Tyr Gly Gln Lys Ser Ala Thr Asn Thr Glu Leu Gln Phe Leu Met Glu
 85 90 95
 Lys Leu Arg Gly Tyr Glu Ser Thr Met Asp Asp Ile Gly Lys Val Tyr
 100 105 110
 Gly Asp Asp Lys Met Arg Asp Ile Ile Lys Asn Ile Ser Asp Asp Asp
 115 120 125
 Ile Lys Ser Leu Leu Gly Glu Ile Asn Ser Asp Tyr Ser Gly Lys Xaa
 130 135 140

(2) INFORMATION FOR SEQ ID NO:12:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 2516 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: SRSV/KY/89

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

CAATAGAGGA TGGCCCTTTA ATTTATGCTG AACATGCCAA GTACAAAAAT CATTTTGATG	60
CAGATTACAC AGCATGGGAC TCTACACAAA ATAGACAAAT TATGACAGAA TCCTTCTCCA	120
TCATGTCACG CCTTACGGCC TCTCCAGAAC TAGCTGAGGT TGTAGCCCAG GACTTACTAG	180
CACCATCCGA GATGGATGTG GCGGACTATG TTATAAGGGT CAAAGAAGGC CTACCATCAG	240
GATTTCCCTG CACTTCTCAA GTGAATAGCA TAAATCACTG GATAATCACC CTTTGTGCAT	300
TGTCTGAGGC TACTGGCTTA TCACCTGATG TGGTACAGTC CATGTCATAC TTCTCATTCT	360
ACGGTGATGA TGAGATCGTA TCAACTGACA TAGACTTTGA CCCAACTCGC CTCACCCAAA	420
TTCTCAAGGA ATACGGCCTC AAGCCAACAA GGCCAGACAA AACAGAAGGA CCAATACAGG	480
TGAGGAAGAA TGTGGATGGG CTAGTTTTTC TCGGCGGCAC CATCTCCCGG GACGCAGCAG	540
GGTTCCAAGG TAGACTGGAT AGAGCCTCAA TTGAACGTCA AATTTTCTGG ACCCGGGGGC	600
CCAAACCATTG AGACCCATCA GAGACTCTGG TACCACACAC CCAAAGGAAA GTCCAGCTGA	660
TCTCACTTCT AGGAGAAGCC TCACTCCACG GGGAAAAATT TTACAGGAAA ATATCTAGCA	720
AAGTCATACA TGAAATTAAG ACTGGTGGGC TGGAGATGTA TGTCCCAGGG TGGCAGGCCA	780
TGTTCCGCTG GATGCGCTTC CATGACCTCG GATTGTGGAC AGGAGATCGC AATCTCCTGC	840
CCGAATTCGT AAATGATGAT GCGCTCTAAG GACGCTACGT CAAGCGTGGA TGGCGCCAGT	900
GCGTCGGTTC AGTTGGTACC GGAGGTTAAT GCTTCTGACC CTCTTGCAAT GGATCCTGTG	960
GCGGGTTCTT CAACAGCAGT TGCAACCGCT GGACAAGTTA ACCCTATTGA CCCTTGATA	1020
ATCAATAACT TTGTGCAGGC TCCCCAAGGT GAATTTACTA TTTCTCCAAA TAATACCCCC	1080
GGTGATGTTT TGTTTGATTG GAGTCTAGGC CCTCATCTTA ATCCCTTCTT GTTACATTTG	1140
TCACAAATGT ATAATGGCTG GGTGGCAAC ATGAGAGTTA GGATTATGCT GGCTGGTAAT	1200
GCATTTACTG CAGGCAAAAT TATAGTTTCT TGCATACCTC CTGGCTTTGG CTCCCAACAA	1260
CTTACTATAG CACAAGCAAC TCTCTTCCCG CATGTGATTG CTGATGTTAG GACTTTAGAC	1320
CCAATTGAAG TACCCTTGGA AGATGTAAGG AATGTTCTCT TTCATAATAA TGATAGAAAT	1380
CAACAAACTA TCGCCTTGT GTGCATGCTT TATACCCCCC TCAGCACTGG TGGCGGTACA	1440
GGTGATTCTT TTGTGGTTGC AGGGCGAGTC ATGACTTGTC CTAGCCCCGA CTTTAATTTT	1500
TTGTCTTGG TTCCTCCAC AGTGGACAG AAGACTAGGC CTTTACCCT CCCAAATTTA	1560
CCGCTGAGTT CTTTGTCTAA TTCACGTGCT CCTCTTCCAA TTAGTGGCAT GGGTATTTCT	1620
CCAGATAATG TTCAGAGTGT GCAGTTCCAA AATGGCCGAT GTACCTTAGA TGGACGTCTT	1680
GTTGGCACCA CCCAGTTTC CCTCTCCCAT GTTGCTAAGA TAAGGGGTAC TTCTAATGGT	1740
ACAGTAATCA ATCTCACCGA ATTGGATGGC ACCCCCTTCC ACCCTTTTGA AGGCCCTGCC	1800
CCTATTGGTT TTCCAGATCT TGGTGGCTGT GATTGGCATA TTAATATGAC ACAATTGGA	1860
CATTCCAGTC AGACTCAGTA TGATGTAGAC ACCACCCCG ACACCTCCGT CCCTCACTTA	1920
GGTTCAATCC AGCGCAATGG CATTGCTAGT GGCAACTATA TTGGTGTCT TAGCTGGGTC	1980

TCCCCCCCAT CACATCCATC TGGCTCTCAA GTTGATCTCT GGAAGATCCC CAACTATGGG	2040
TCTAGTATCA CAGAGGCAAC CCATCTAGCT CCCTCTGTCT ATTCTCCTGG CTTTGGAGAG	2100
GTGCTAGTCT TTTTCATGTC AAAGATACCA GGTCTGCTG GTGATAGTCT GCCCTGTTTA	2160
CTGCCACAAG GATATATCTC ACACCTTGCA AGTGAACAAG CCCCAACTGT TGGTGAGGGT	2220
CCCCTGCTCC ACTATGTTGA CCCTGACAG GACCGGAATC TTGGGGAGTT TAAGGCTTAC	2280
CCTGATGGTT TCCTAACCTG TGTCCCTAAT GGGGCCAGCT CGGGCCCACA ACAACTACCA	2340
ATCAATGGAG TCTTTGTCTT TGTTCATGG GTGTCCAGAT TTTATCAGTT AAAGCCTGTG	2400
GGAAGTCCCA GTACGGCAAG AGGTAGGCTT GGTTCGCGCC GATAATGGCT CAGGCTATAA	2460
TTGGTGCAAT TGCCGCCTCT ACAGCAGGTA GTGCTTAGG GGCAGGTATA CAGGTT	2516

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 124 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: primate calcicivirus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TGGACGGACC TGCTGTTGAA GATCTCTTCA AAGGCTCGAA CGACCAAAGC ACGATCGGTA	60
TTGTGTTGAC TACGCAAAGT GGGACTCAAC CCACCACCAA AAGTAACATC CAATCAATGA	120
CATC	124

(2) INFORMATION FOR SEQ ID NO:14:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 110 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: primate calcicivirus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:14:

GTGAATGACA TCTTCGACTC GATGGACCTA TTCACATATG GTGATGACGG TGTCTACATC	60
GTCCCACCAC TATATCATCT GTCATGCCCA AGTCTTCACC AACCTGAAAC	110

(2) INFORMATION FOR SEQ ID NO:15:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 21 base pairs

(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:15:

CTTGTTGGTT TGAGGCCATA T

21

(2) INFORMATION FOR SEQ ID NO:16:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 19 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:16:

ATAAAAGTTG GCATGAACA

19

(2) INFORMATION FOR SEQ ID NO:17:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:17:

GTTGACACAA TCTCATCATC

20

(2) INFORMATION FOR SEQ ID NO:18:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:18:

GGCCTGCCAT CTGGATTGCC

20

(2) INFORMATION FOR SEQ ID NO:19:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 20 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:19:

GGGCCCCCTG GTATAGGTAA

20

(2) INFORMATION FOR SEQ ID NO:20:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 28 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:20:

TGGTGATGAC TATAGCATCA GACACAAA

28

(2) INFORMATION FOR SEQ ID NO:21:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:21:

ACTCACCCAA ATCCTCCA

18

(2) INFORMATION FOR SEQ ID NO:22:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: unknown
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:22:

GTTCTGACCA CCTAACCT

18

(2) INFORMATION FOR SEQ ID NO:23:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 26 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:23:

AGTTTGGGTC CCCATCTTAA TCCTTT

26

(2) INFORMATION FOR SEQ ID NO:24:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:24:

TGAACCAAAA CCAGGGGG

18

(2) INFORMATION FOR SEQ ID NO:25:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:25:

AGCAAAGTCA TACATGAAAT

20

(2) INFORMATION FOR SEQ ID NO:26:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 17 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:26:

CCATTATACA TTTGTAG

17

(2) INFORMATION FOR SEQ ID NO:27:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

- (ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:27:

ATTATAGTTT CTTGCATA

18

(2) INFORMATION FOR SEQ ID NO:28:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid

- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:28:

CACACTCTGG ACATTGTCTG

20

(2) INFORMATION FOR SEQ ID NO:29:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 19 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:29:

CATTGGGTTT CCAGACCTA

19

(2) INFORMATION FOR SEQ ID NO:30:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 20 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:30:

ATAATTGGGG ATCTTCAAA

20

(2) INFORMATION FOR SEQ ID NO:31:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:31:

TAGTGGCATG GGTATTC

18

(2) INFORMATION FOR SEQ ID NO:32:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:32:

TATGCCAATC ACAGCCAC

18

(2) INFORMATION FOR SEQ ID NO:33:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 20 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:33:

GTCTGGCTCC CAAGTTGACC

20

(2) INFORMATION FOR SEQ ID NO:34:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:34:

CGGTATCAGG GTCAACAT

18

(2) INFORMATION FOR SEQ ID NO:35:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:35:

TGAGGCTGCC CTGCTCCA

18

(2) INFORMATION FOR SEQ ID NO:36:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 18 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:36:

CCACCGCTGT CCGGGAGG

18

(2) INFORMATION FOR SEQ ID NO:37:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 19 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:37:

GTTGCTGTTG GCATTACA

19

(2) INFORMATION FOR SEQ ID NO:38:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 126 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE
 (A) ORGANISM: Norwalk virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:38:

His	Phe	Asp	Ala	Asp	Tyr	Thr	Ala	Trp	Asp	Ser	Thr	Gln	Asn	Arg	Gln	1	5	10	15
Ile	Met	Thr	Glu	Ser	Phe	Ser	Ile	Met	Ser	Arg	Leu	Thr	Ala	Ser	Pro	20	25	30	
Glu	Leu	Ala	Glu	Val	Val	Ala	Gln	Asp	Leu	Leu	Ala	Pro	Ser	Glu	Met	35	40	45	
Asp	Val	Gly	Asp	Tyr	Val	Ile	Arg	Val	Lys	Glu	Gly	Pro	Ser	Gly	Phe	50	55	60	
Pro	Cys	Thr	Ser	Gln	Val	Asn	Ser	Ile	Asn	His	Trp	Ile	Ile	Thr	Leu	65	70	75	80
Cys	Ala	Leu	Ser	Glu	Ala	Thr	Gly	Leu	Ser	Pro	Asp	Val	Val	Gln	Ser	85	90	95	
Met	Ser	Tyr	Phe	Ser	Phe	Tyr	Gly	Asp	Asp	Glu	Ile	Val	Ser	Thr	Asp	100	105	110	
Ile	Asp	Phe	Asp	Pro	Ala	Arg	Leu	Thr	Gln	Ile	Leu	Lys	Glu	115	120	125			

(2) INFORMATION FOR SEQ ID NO:39:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 121 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:
 (A) ORGANISM: hepatitis E virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:39:

Val Phe Glu Asn Asp Phe Ser Glu Phe Asp Ser Thr Gln Asn Asn Phe
 1 5 10 15
 Ser Leu Gly Leu Glu Cys Ala Ile Met Glu Glu Cys Gly Met Pro Gln
 20 25 30
 Trp Leu Ile Arg Leu Tyr His Leu Ile Arg Ser Ala Trp Ile Leu Gln
 35 40 45
 Ala Pro Lys Glu Ser Leu Arg Gly Phe Trp Lys Lys His Ser Lys His
 50 55 60
 Ser Gly Glu Pro Gly Thr Leu Leu Trp Asn Thr Val Trp Asn Met Ala
 65 70 75 80
 Val Ile Thr His Cys Tyr Asp Phe Arg Asp Phe Gln Val Ala Ala Phe
 85 90 95
 Lys Gly Asp Asp Ser Ile Val Leu Cys Ser Glu Tyr Arg Gln Ser Pro
 100 105 110
 Gly Ala Ala Val Leu Ile Ala Gly Cys
 115 120

(2) INFORMATION FOR SEQ ID NO:40:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 127 amino acids
 (B) TYPE: amino acid
 (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: hepatitis C virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:40:

Gly Phe Ser Tyr Asp Thr Arg Cys Phe Asp Ser Thr Val Thr Glu Ser
 1 5 10 15
 Asp Ile Arg Thr Glu Glu Ala Ile Tyr Gln Cys Cys Asp Leu Asp Pro
 20 25 30
 Gln Ala Arg Val Ala Ile Lys Ser Leu Thr Glu Arg Leu Tyr Val Gly
 35 40 45
 Gly Pro Leu Thr Asn Ser Arg Gly Glu Asn Cys Gly Tyr Arg Arg Cys
 50 55 60
 Arg Ala Ser Arg Ala Ser Gly Val Leu Thr Thr Ser Cys Gly Asn Thr
 65 70 75 80
 Leu Thr Cys Tyr Ile Lys Ala Arg Ala Ala Cys Arg Ala Ala Gly Leu
 85 90 95
 Gln Asp Cys Thr Met Leu Val Cys Gly Asp Asp Leu Val Val Ile Cys
 100 105 110
 Glu Ser Ala Gly Val Gln Glu Asp Ala Ala Ser Leu Arg Ala Phe
 115 120 125

(2) INFORMATION FOR SEQ ID NO:41:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 132 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: hepatitis A virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:41:

Gly Leu Asp Leu Asp Phe Ser Ala Phe Asp Ala Ser Leu Ser Pro Phe
 1 5 10 15
 Met Ile Arg Glu Ala Gly Arg Ile Met Ser Glu Leu Ser Gly Thr Pro
 20 25 30
 Ser His Phe Gly Thr Ala Leu Ile Asn Thr Ile Ile Tyr Ser Lys His
 35 40 45
 Leu Leu Tyr Asn Cys Cys Tyr His Val Cys Gly Ser Met Pro Ser Gly
 50 55 60
 Ser Pro Cys Thr Ala Leu Leu Asn Ser Ile Ile Asn Asn Val Asn Leu
 65 70 75 80
 Tyr Tyr Val Phe Ser Lys Ile Phe Gly Lys Ser Pro Val Phe Phe Cys
 85 90 95
 Gln Ala Leu Lys Ile Leu Cys Tyr Gly Asp Asp Val Leu Ile Val Phe
 100 105 110
 Ser Arg Asp Val Gln Ile Asp Asn Leu Asp Leu Ile Gly Gln Lys Ile
 115 120 125
 Val Asp Glu Phe
 130

(2) INFORMATION FOR SEQ ID NO:42:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 158 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Japanese encephalitis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:42:

Met Tyr Ala Asp Asp Thr Ala Gly Trp Asp Thr Arg Ile Thr Arg Thr
 1 5 10 15
 Asp Leu Glu Asn Glu Ala Lys Val Leu Glu Leu Leu Asp Gly Glu His
 20 25 30
 Arg Met Leu Ala Arg Ala Ile Ile Glu Leu Thr Tyr Arg His Lys Val
 35 40 45

85

Val Lys Val Met Arg Pro Ala Ala Glu Gly Lys Thr Val Met Asp Val
 50 55 60

Ile Ser Arg Glu Asp Gln Arg Gly Ser Gly Gln Val Val Thr Tyr Ala
 65 70 75 80

Leu Asn Thr Phe Thr Asn Ile Ala Val Gln Leu Val Arg Leu Met Glu
 85 90 95

Ala Glu Gly Val Ile Gly Pro Gln His Leu Glu Gln Leu Pro Arg Lys
 100 105 110

Thr Lys Ile Ala Val Arg Thr Trp Leu Phe Glu Asn Gly Glu Glu Arg
 115 120 125

Val Thr Arg Met Ala Ile Ser Gly Asp Asp Cys Val Val Lys Pro Leu
 130 135 140

Asp Asp Arg Phe Ala Thr Ala Leu His Phe Leu Asn Ala Met
 145 150 155

(2) INFORMATION FOR SEQ ID NO:43:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 117 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Poliovirus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:43:

Phe Ala Phe Asp Tyr Thr Gly Tyr Asp Ala Ser Leu Ser Pro Ala Trp
 1 5 10 15

Phe Glu Ala Leu Lys Met Val Leu Glu Lys Ile Gly Phe Gly Asp Arg
 20 25 30

Val Asp Tyr Ile Asp Tyr Leu Asn His Ser His His Leu Tyr Lys Asn
 35 40 45

Lys Thr Tyr Cys Val Lys Gly Gly Met Pro Ser Gly Cys Ser Gly Thr
 50 55 60

Ser Ile Phe Asn Ser Met Ile Asn Asn Leu Ile Ile Arg Thr Leu Leu
 65 70 75 80

Leu Lys Thr Tyr Lys Gly Ile Asp Leu Asp His Leu Lys Met Ile Ala
 85 90 95

Tyr Gly Asp Asp Val Ile Ala Ser Tyr Pro His Glu Val Asp Ala Ser
 100 105 110

Leu Leu Ala Gln Ser
 115

(2) INFORMATION FOR SEQ ID NO:44:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 121 amino acids

(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Foot-and-mouth disease virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:44:

```

Val Trp Asp Val Asp Tyr Ser Ala Phe Asp Ala Asn His Cys Ser Asp
 1           5           10           15
Ala Met Asn Ile Met Phe Glu Glu Val Phe Arg Thr Asp Phe Gly Phe
 20           25           30
His Pro Asn Ala Glu Trp Ile Leu Lys Thr Leu Val Asn Thr Glu His
 35           40           45
Ala Tyr Glu Asn Lys Arg Ile Thr Val Glu Gly Gly Met Pro Ser Gly
 50           55           60
Cys Ser Ala Thr Ser Ile Ile Asn Thr Ile Leu Asn Asn Ile Tyr Val
 65           70           75           80
Leu Tyr Ala Leu Arg Arg His Tyr Glu Gly Val Glu Leu Asp Thr Tyr
 85           90           95
Thr Met Ile Ser Tyr Gly Asp Asp Ile Val Val Ala Ser Asp Tyr Asp
 100          105          110
Leu Asp Phe Glu Ala Leu Lys Pro His
 115          120

```

(2) INFORMATION FOR SEQ ID NO:45:

(i) SEQUENCE CHARACTERISTICS:

(A) LENGTH: 126 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

(A) ORGANISM: encephalomyocarditis virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:45:

```

Val Tyr Asp Val Asp Tyr Ser Asn Phe Asp Ser Thr His Ser Val Ala
 1           5           10           15
Met Phe Arg Leu Leu Ala Glu Glu Phe Phe Thr Pro Glu Asn Gly Phe
 20           25           30
Asp Pro Leu Thr Arg Glu Tyr Leu Glu Ser Leu Ala Ile Ser Thr His
 35           40           45
Ala Phe Glu Glu Lys Arg Phe Leu Ile Thr Gly Gly Leu Pro Ser Gly
 50           55           60
Cys Ala Ala Thr Ser Met Leu Asn Thr Ile Met Asn Asn Ile Ile Ile
 65           70           75           80

```


(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 122 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:
(A) ORGANISM: Sindbis virus

Val 1	Leu	Glu	Thr	Asp 5	Ile	Ala	Ser	Phe	Asp 10	Lys	Ser	Gln	Asp 15	Ala
Met	Ala	Leu	Thr 20	Gly	Leu	Met	Ile	Leu 25	Glu	Asp	Leu	Gly	Val 30	Asp Gln
Pro	Leu	Leu 35	Asp	Leu	Ile	Glu	Cys 40	Ala	Phe	Gly	Glu	Ile 45	Ser	Ser Thr
His 50	Leu	Pro	Thr	Gly	Thr	Arg 55	Phe	Lys	Phe	Gly	Ala 60	Met	Met	Lys Ser
Gly 65	Met	Phe	Leu	Thr 70	Leu	Phe	Val	Asn	Thr	Val 75	Leu	Asn	Val	Val Ile 80
Ala	Ser	Arg	Val 85	Leu	Glu	Glu	Arg	Leu	Lys 90	Thr	Ser	Arg	Cys	Ala Ala 95
Phe	Ile	Gly	Asp 100	Asp	Asn	Ile	Ile	His 105	Gly	Val	Val	Ser	Asp 110	Lys Glu
Met	Ala	Glu 115	Arg	Cys	Ala	Thr	Trp 120	Leu	Asn					

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 124 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(vi) ORIGINAL SOURCE:
(A) ORGANISM: tobacco mosaic virus

Val Leu Glu Leu Asp Ile Ser Lys Tyr Asp Lys Ser Gln Asn Glu Phe
1 5 10 15

```

His Cys Ala Val Glu Tyr Glu Ile Trp Arg Arg Leu Gly Phe Glu Asp
      20                      25                      30
Phe Leu Gly Glu Val Trp Lys Gln Gly His Arg Lys Thr Thr Leu Lys
      35                      40                      45
Asp Ile Thr Ala Gly Tyr Lys Thr Cys Ile Trp Tyr Gln Arg Lys Ser
      50                      55                      60
Gly Asp Val Thr Thr Phe Ile Gly Asn Thr Val Ile Ile Ala Ala Cys
      65                      70                      75                      80
Leu Ala Ser Met Leu Pro Met Glu Lys Ile Ile Lys Gly Ala Phe Cys
      85                      90                      95
Gly Asp Asp Ser Leu Leu Tyr Phe Pro Lys Gly Cys Glu Phe Pro Asp
      100                     105                     110
Val Gln His Ser Ala Asn Leu Met Trp Asn Phe Glu
      115                     120

```

(2) INFORMATION FOR SEQ ID NO:48:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 125 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: peptide
- (vi) ORIGINAL SOURCE:
 - (A) ORGANISM: alfalfa mosaic virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:48:

```

Phe Lys Glu Ile Asp Phe Ser Lys Phe Asp Lys Ser Gln Asn Glu Leu
1      5      10      15
His His Leu Ile Gln Glu Arg Phe Leu Lys Tyr Leu Gly Ile Pro Asn
      20      25      30
Glu Phe Leu Thr Leu Trp Phe Asn Ala His Arg Lys Ser Arg Ile Ser
      35      40      45
Asp Ser Lys Asn Gly Val Phe Phe Asn Val Asp Phe Gln Arg Arg Thr
      50      55      60
Gly Asp Ala Leu Thr Tyr Leu Gly Asn Thr Ile Val Thr Leu Ala Cys
      65      70      75      80
Leu Cys His Val Tyr Asp Leu Met Asp Pro Asn Val Lys Phe Val Val
      85      90      95
Ala Ser Gly Asp Asp Ser Leu Ile Gly Thr Val Glu Glu Leu Pro Arg
      100     105     110
Asp Gln Glu Phe Leu Phe Thr Thr Leu Phe Asn Leu Glu
      115     120     125

```

(2) INFORMATION FOR SEQ ID NO:49:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 122 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: brome mosaic virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:49:

```

Phe Leu Glu Ala Asp Leu Ser Lys Phe Asp Lys Ser Gln Gly Glu Leu
1           5           10           15
His Leu Glu Phe Gln Arg Glu Ile Leu Leu Ala Leu Gly Phe Pro Ala
20          25          30
Pro Leu Thr Asn Trp Trp Ser Asp Phe His Arg Asp Ser Tyr Leu Ser
35          40          45
Asp Pro His Ala Lys Val Gly Met Ser Val Ser Phe Gln Arg Arg Thr
50          55          60
Gly Asp Ala Phe Thr Tyr Phe Gly Asn Thr Leu Val Thr Met Ala Met
65          70          75          80
Ile Ala Tyr Ala Ser Asp Leu Ser Asp Cys Asp Cys Ala Ile Phe Ser
85          90          95
Gly Asp Asp Ser Leu Ile Ile Ser Lys Val Lys Pro Val Leu Asp Thr
100         105         110
Asp Met Phe Thr Ser Leu Phe Asn Met Glu
115         120

```

(2) INFORMATION FOR SEQ ID NO:50:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 142 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: peptide

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: cowpea mosaic virus

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:50:

```

Val Leu Cys Cys Asp Tyr Ser Ser Phe Asp Gly Leu Leu Ser Lys Gln
1           5           10           15
Val Met Asp Val Ile Ala Ser Met Ile Asn Glu Leu Cys Gly Gly Glu
20          25          30
Asp Gln Leu Lys Asn Ala Arg Arg Asn Leu Leu Met Ala Cys Cys Ser
35          40          45
Arg Leu Ala Ile Cys Lys Asn Thr Val Trp Arg Val Glu Cys Gly Ile
50          55          60

```

90

Pro Ser Gly Phe Pro Met Thr Val Ile Val Asn Ser Ile Phe Asn Glu
 65 70 75 80

Ile Leu Ile Arg Tyr His Tyr Lys Lys Leu Met Arg Glu Gln Gln Ala
 85 90 95

Pro Glu Leu Met Val Gln Ser Phe Asp Lys Leu Ile Gly Leu Val Thr
 100 105 110

Tyr Gly Asp Asp Asn Leu Ile Ser Val Asn Ala Val Val Thr Pro Tyr
 115 120 125

Phe Asp Gly Lys Lys Leu Lys Gln Ser Leu Ala Gln Gly Gly
 130 135 140

(2) INFORMATION FOR SEQ ID NO:51:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:51:

CACGCGGAGG CTCTCAAT

18

(2) INFORMATION FOR SEQ ID NO:52:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:52:

GGTGGCGAAG CGGCCCTC

18

(2) INFORMATION FOR SEQ ID NO:53:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 base pairs
 (B) TYPE: nucleic acid
 (C) STRANDEDNESS: double
 (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:53:

TCAGCAGTTA TAGATATG

18

(2) INFORMATION FOR SEQ ID NO:54:

- (i) SEQUENCE CHARACTERISTICS:
 (A) LENGTH: 18 base pairs

- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:54:

ATGCTATATA CATAGGTC

18

(2) INFORMATION FOR SEQ ID NO:55:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:55:

CAACAGGTAC TACGTGAC

18

(2) INFORMATION FOR SEQ ID NO:56:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 18 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:56:

TGTGGCCCAA GATTGCT

18

(2) INFORMATION FOR SEQ ID NO:57:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:57:

ATAAAAGTTG GCATGAACAC AAAT

24

(2) INFORMATION FOR SEQ ID NO:58:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 24 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: double
 - (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:58:
GTTGCTGTTG GCATTAAACAT GGAC 24
- (2) INFORMATION FOR SEQ ID NO:59:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:59:
GTTCTGTTG GCATAAACAT GGAC 24
- (2) INFORMATION FOR SEQ ID NO:60:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:60:
GTTCCGGTTG GCATTAAACAT GGAC 24
- (2) INFORMATION FOR SEQ ID NO:61:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:61:
GTTCCGGTTG GTATCAACAT GGAC 24
- (2) INFORMATION FOR SEQ ID NO:62:
- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 24 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown
- (ii) MOLECULE TYPE: DNA (genomic)
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:62:
GTTCCGGTTG GTGTTGACAT GACA 24

(2) INFORMATION FOR SEQ ID NO:63:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 118 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: SRSV/CDC 6/91

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:63:

ATGCACTTCA CAGGTGAATA GCATCAACCA CTGGATCCTA ACTCTATGTG CATTGTCAGA 60
AGTCACTGGC TTGTCCCCTG ATGTGATACA ATCACAATCT TATTTCTCAT TTTATGGT 118

(2) INFORMATION FOR SEQ ID NO:64:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 118 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: SRSV/UT/88

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:64:

ATGTACCTCA CAAGTGAACA GCATCAATCA CTGGATTTTG ACCTTGTGGG GCCTATCAGA 60
AGTTACTGGT CTGGCTCCTG ATGTAATACA GTCACAATCT TACTTTTCAT TCTATGGT 118

(2) INFORMATION FOR SEQ ID NO:65:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 117 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: double
(D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

- (vi) ORIGINAL SOURCE:
(A) ORGANISM: Snow Mountain Agent/78

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:65:

CTGCACATCA CAGTGGGAATT CCATGCCAC TGGCTCCTCA CACTCTGTGC ACTATCTGAA 60
GTCACAAACC TGGCTCCTGA CATCATACAA GCTAACTCCT TGTTCTCTTT CTATGGT 117

(2) INFORMATION FOR SEQ ID NO:66:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 118 base pairs
(B) TYPE: nucleic acid

- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: SRSV/CAMBRIDGE, UK 92

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:66:

```
CTGCACCTCA CAGTGGAACCT CCATTGCCCA CTGGTTGCTT ACTCTGTGTG CCCTTTCTGA      60
AGTGACAGGA CTAGGCCCCG ACATCATACA AGCTAATTCC ATGTACTCTT TCTATGGT      118
```

(2) INFORMATION FOR SEQ ID NO:67:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: SRSV/CDC 32

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:67:

```
TTGCACCTCA CAGTGGAACCT CCATTGCCCT CTGGTTGCTT ACTCTGTGTG CCCTTTCTGA      60
AGTGACAGGA CTAGGCCCCG ACATCATACA AGCTAATTCC ATGTACTCTT TCTATGGT      118
```

(2) INFORMATION FOR SEQ ID NO:68:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Norwalk virus/8FIIa/68

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:68:

```
ATGTACTTCC CAGGTGAACA GCATAAATCA CTGGATAATT ACTCTCTGTG CACTGTCTGA      60
GGCCACTGGT TTATCACCTG ATGTGGTGCA ATCCATGTCA TATTTCTCAT TTTATGGT      118
```

(2) INFORMATION FOR SEQ ID NO:69:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: SRSV-3/88

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:69:

CTGCACTTCT CAAGTAAATA GCATAAATCA CTGGATAATC ACCCTTTGTG CACTGTCTGA 60
GGCTACTGGC TTATCACCTG ATGTGGTGCA GTCCATGTCA TACTTCTCAT TTTACGGT 118

(2) INFORMATION FOR SEQ ID NO:70:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 118 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: SRSV/KY89/89

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:70:

CTGCACTTCT CAAGTGAATS GCATAAATCA CTGGATAATC ACCCTTTGTG CATTGTCTGA 60
GGCTACTGGC TTATCACCTG ATGTGGTACA GTCCATGTCA TACTTCTCAT TCTACGGT 118

(2) INFORMATION FOR SEQ ID NO:71:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 279 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: Norwalk Virus/8FIIa/68

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:71:

CAATAGAAGA TGGCCCCCTC ATCTATGCTG AGCATGCTAA ATATAAGAAT CATTTTGATG 60
CAGATTATAC AGCATGGGAC TCAACACAAA ATAGACAAAT TATGACAGAA TCCTTCTCCA 120
TTATGTGCGC CCTTACGGCC TCACCAGAAT TGGCCGAGGT TGTGGCCCAA GATTTGCTAG 180
CACCATCTGA GATGGATGTA GGTGATTATG TCATCAGGGT CAAAGAGGGG CTGCCATCTG 240
GATTCCCATG TACTTCCCAG GTGAACAGCA TAAATCACT 279

(2) INFORMATION FOR SEQ ID NO:72:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 279 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: SRSV-3/88

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:72:

CAATAGAGGA TGGCCCTTTA ATTTATGCTG AGCATGCCAA GTACAAAAAT CATTGTGATG	60
CAGATTACAC AGCATGGGAC TCTACACAAA ATAGACAAAT AATGACAGAA TCCTTTTCCA	120
TCATGTCACG CCTCACGGCC TCTCCAGAAC TAGCTGAGGT TGTAGCCCAG GACTTGCTAG	180
CACCATCCGA GATGGATGTG GGTGACTATG TTATAAGGGT CAAAGAAGGC CTACCATCAG	240
GATTTCCTG CACTTCTCAA GTAAATAGCA TAAATCACT	279

(2) INFORMATION FOR SEQ ID NO:73:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 279 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: SRSV/KY89

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:73:

CAATAGAGGA TGGCCCTTTA ATTTATGCTG AACATGCCAA GTACAAAAAT CATTGTGATG	60
CAGATTACAC AGCATGGGAC TCTACACAAA ATAGACAAAT TATGACAGAA TCCTTCTCCA	120
TCATGTCACG CCTTACGGCC TCTCCAGAAC TAGCTGAGGT TGTAGCCCAG GACTTACTAG	180
CACCATCCGA GATGGATGTG GCGGACTATG TTATAAGGGT CAAAGAAGGC CTACCATCAG	240
GATTTCCTG CACTTCTCAA GTGAATAGCA TAAATCACT	279

(2) INFORMATION FOR SEQ ID NO:74:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 279 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

(A) ORGANISM: SRSV/Cambridge, UK/92

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:74:

TGTATGAAGA TGGTACCATA ATATTTGAGA AACATTCCAG ATACAGATAC CACTATGATG	60
CAGATTATCC CGCTGGGTAC TCCACGCAGC AACGGGCAGT GTTGGCAGCA GCACTTGAAA	120
TCATGGTGAG GTTCTCTGCT GAACACAGC TAGCGCAAAT AGTAGCTGAA GATCTGCTAG	180
CACCAAGTGT AGTTGATGTG GGTGACTTCA AGATCACCAT TAATGAAGGC CTACCTTCTG	240

GTGTGCCCTG CACCTCACAG TGGAACCTCCA TTGCCCACT

279

(2) INFORMATION FOR SEQ ID NO:75:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 277 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: double
- (D) TOPOLOGY: unknown

(ii) MOLECULE TYPE: DNA (genomic)

(vi) ORIGINAL SOURCE:

- (A) ORGANISM: Snow Mountain Agent/78

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:75:

GAATGAGGAT GGACCCATAA TTTTGTAAAA GCACTCCAGG TTCTCATACC ACTATGATGC	60
AGATTACTCA CGCTGGGACT CAACCCAACA GAGGGCAGTG CTAGCTGCAG CCTTGGAAT	120
CATGGTAAAA TTCTCACCAG AACCACATTT GGCCCAAATT GTTGCAGAGG ATCTCCTAGC	180
CCCCAGTGTG ATGGATGTAG GTGATTTC AAATAACAATT AATGAGGGAC TGCCCTCGGG	240
AGTACCCTGC ACATCACAGT GGAATTCAT GCCCACT	277

CLAIMS

1. A cDNA sequence of the formula shown in Table 2 and fragments and derivatives thereof having sufficient size to bind a Norwalk or Norwalk-related virus genome.
- 5 2. A protein encoded by nucleotides including nucleotides 1 through 7753 of the Norwalk virus genome shown in Table 2 or fragments or derivatives thereof.
3. The protein of claim 2, wherein said protein is produced in a prokaryotic expression system or a eukaryotic expression system.
- 10 4. The protein of claim 2, wherein said protein is produced by chemical methods.
5. A protein encoded by nucleotides 146 through 5359 of the Norwalk virus genome shown in Table 2 or fragments or derivatives thereof.
- 15 6. The protein of claim 5, wherein said protein is produced in a prokaryotic expression system or eukaryotic expression system.
7. The protein of claim 5, wherein said protein is produced by chemical methods.
8. A RNA-dependent RNA polymerase encoded by nucleotides
20 4543 to 4924 of the Norwalk virus genome shown in Table 2 or fragments.
9. The RNA polymerase of claim 8, wherein said RNA polymerase is produced in a prokaryotic expression system or a eukaryotic expression system.

10. The RNA polymerase of claim 8, wherein said RNA polymerase is produced by chemical methods.

11. A protein encoded by nucleotides 5337 through 7573 of the Norwalk virus genome shown in Table 2 or fragments or derivatives thereof.

12. The protein of claim 11, wherein said protein is produced in a prokaryotic expression system or eukaryotic expression system.

13. The protein of claim 11, wherein said protein is produced by chemical methods.

14. A protein encoded by nucleotides 5346 through 6935 of the Norwalk virus genome shown in Table 2 or fragments or derivatives thereof.

15. The protein of claim 14, wherein said protein is produced in a prokaryotic expression system or eukaryotic expression system.

16. The protein of claim 14, wherein said protein is produced by chemical methods.

17. A protein encoded by nucleotides 6938 through 7573 of the Norwalk virus genome shown in Table 2 or fragments or derivatives thereof.

18. The protein of claim 17, wherein said protein is produced in a prokaryotic expression system or eukaryotic expression system.

19. The protein of claim 17, wherein said protein is produced by chemical methods.

20. A method of making a RNA probe to detect Norwalk or Norwalk-related viruses, comprising the steps of:

subcloning a Norwalk virus cDNA clone into a transcription vector;

5 growing said cDNA containing transcription vector;
adding RNA polymerase to generate single stranded RNA by

in vitro transcription; and

isolating said single stranded RNA.

21. A method of identifying Norwalk or Norwalk-related viruses
10 in a sample suspected of containing Norwalk or Norwalk-related viruses,
comprising the steps of:

adding a cDNA or a RNA probe specific to Norwalk virus or
a Norwalk-related virus to said sample to be tested under
conditions in which the cDNA or RNA probe will bind to the
15 Norwalk or Norwalk-related virus genome; and

measuring the amount of binding of said cDNA or RNA
probe.

22. The method of claim 21, wherein said sample is selected from
the group consisting of food, water and stool.

20 23. The method of claim 21, wherein said cDNA is selected from
a group consisting of pUCNV-953, pUCNV-4145, pUCNV-4095, pUCNV-
5030 and pUCNV-5101 or fragments or derivatives thereof.

24. A method of identifying Norwalk or Norwalk-related viruses
in a sample suspected of containing Norwalk or Norwalk-related viruses
25 comprising the steps of:

adding at least two oligonucleotides each of about 10
nucleotides or greater to said sample under conditions in which
said oligonucleotides bind to the Norwalk or Norwalk-related virus
genome;

amplifying a nucleotide sequence between said bound oligonucleotides; and
measuring the amount of amplified sequence.

25. A method of identifying Norwalk or Norwalk-related viruses
5 in a sample suspected of containing Norwalk or Norwalk-related viruses comprising the steps of:

isolating said nucleic acids using CTAB procedure;
amplifying nucleic acid; and
measuring the amplified product.

10 26. The method of claim 25, wherein the CTAB procedure includes:

extracting said sample with genetron;
removing the supernatant of said genetron extracted
sampled;

15 precipitating viruses in said supernatant with polyethylene glycol;

treating said precipitate with proteinase K in the presence of SDS at about 30° minutes;

20 sequentially extracting said treated precipitate with phenol-chloroform and then chloroform;

forming a mixture by adding a solution of about 5% CTAB and about 0.4M NaCl to said supernatant of said sequentially extracted sample at a ratio of about 5:2 sample:CTAB;

25 incubating said mixture;
centrifuging said mixture to collect nucleic acids;
suspending said nucleic acids in 1M NaCl and thereafter extracting with chloroform.

27. A method of claim 25 further comprising:
performing reverse transcription on said nucleic acids;
30 amplifying nucleic acids using primers; and

detecting the amplified nucleic acids using agarose gel electrophoresis.

28. A method of cloning Norwalk or pathogens from food, biological and environmental samples, comprising:
- 5 isolating said nucleic acids using CTAB procedure;
amplifying nucleic acids; and
incorporating said amplified nucleic acids into vectors.
29. A primer sequence of the formula CTT GTT GGT TTG AGG
CCA TAT.
- 10 30. A primer sequence of the formula ATA AAA GTT GGC ATG
AAC A.
31. A primer sequence of the formula GTT GAC ACA ATC TCA
TCA TC.
32. A primer sequence of the formula GGC CTG CCA TCT GGA
15 TTG CC.
33. A primer sequence of the formula GGG CCC CCT GGT ATA
GGT AA.
34. A primer sequence of the formula TGG TGA TGA CTA TAG
CAT CAG ACA CAA A.
- 20 35. A primer sequence of the formula ACT CAC CCA AAT CCT
CCA.
36. A primer sequence of the formula GTT CTG ACC ACC TAA
CCT.

37. A primer sequence of the formula AGT TTG GGT CCC CAT
CTT AAT CCT TT.
38. A primer sequence of the formula TGA ACC AAA ACC AGG
GGG.
- 5 39. A primer sequence of the formula AGC AAA GTC ATA CAT
GAA AT.
40. A primer sequence of the formula CCA TTA TAC ATT TGT
AG.
41. A primer sequence of the formula ATT ATA GTT TCT TGC
10 ATA.
42. A primer sequence of the formula CAC ACT CTG GAC ATT
GTC TG.
43. A primer sequence of the formula CAT TGG GTT TCC AGA
CCT A.
- 15 44. A primer sequence of the formula ATA ATT GGG GAT CTT
CCA AA.
45. A primer sequence of the formula TAG TGG CAT GGG TAT
TTC.
46. A primer sequence of the formula TAT GCC AAT CAC AGC
20 CAC.
47. A primer sequence of the formula GTC TGG CTC CCA AGT
TGA CC.

48. A primer sequence of the formula CGG TAT CAG GGT CAA
CAT.
49. A primer sequence of the formula TGA GGC TGC CCT GCT
CCA.
- 5 50. A primer sequence of the formula CCA CCG CTG TCC GGG
AGG.
51. A primer sequence of the formula GTT GCT GTT GGC ATT
AAC A.
52. A method of making a probe to detect Norwalk or Norwalk-
10 related viruses, comprising the steps of:
synthesizing one or more short or long nucleotides from the
Norwalk virus genome shown in Table 2 or fragments or
derivatives thereof.
53. The probe produced by the method of claim 52.
- 15 54. A method of making a probe to detect Norwalk or Norwalk-
related viruses, comprising the step of:
synthesizing one or more short or long nucleotides from a
subgenomic region of the Norwalk virus genome shown in Table 2
or fragments or derivatives thereof.
- 20 55. The probe produced by the method of claim 54.
56. The probe of claim 55, wherein said subgenomic region
includes a sequence of the formula CTT GTT GGT TTG AGG CCA TAT.

57. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula ATA AAA GTT GGC ATG AAC A.

58. The probe of claim 55, wherein said subgenomic region
5 includes a nucleotide sequence of the formula GTT GAC ACA ATC TCA TCA TC.

59. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula GGC CTG CCA TCT GGA TTG CC.

10 60. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula GGG CCC CCT GGT ATA GGT AA.

61. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula TGG TGA TGA CTA TAG
15 CAT CAG ACA CAA A.

62. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula GTT CTG ACC ACC TAA CCT.

63. The probe of claim 55, wherein said subgenomic region
20 includes a nucleotide sequence of the formula AGT TTG GGT CCC CAT CTT AAT CCT TT.

64. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula TGA ACC AAA ACC AGG GGG.

65. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula AGC AAA GTC ATA CAT GAA AT.

5 66. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula CCA TTA TAC ATT TGT AG.

67. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula CAC ACT CTG GAC ATT GTC TG.

10 68. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula CAT TGG GTT TCC AGA CCT A.

69. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula ATA ATT GGG GAT CTT
15 CCA AA.

70. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula TAT GCC AAT CAC AGC CAC.

71. The probe of claim 55, wherein said subgenomic region
20 includes a nucleotide sequence of the formula GTC TGG CTC CCA AGT TGA CC.

72. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula CGG TAT CAG GGT CAA CAT.

73. The probe of claim 55, wherein said subgenomic region includes a nucleotide sequence of the formula TGA GGC TGC CCT GCT CCA.

74. The probe of claim 55, wherein said subgenomic region
5 includes a nucleotide sequence of the formula CCA CCG CTG TCC GGG AGG.

75. The method of claim 54, wherein said subgenomic region includes said Norwalk genome's first open reading frame.

76. The probe produced by the method of claim 75.

10 77. The method of claim 54, wherein said subgenomic region includes nucleotides 146 through 5359.

78. The probe produced by the method of claim 77.

79. The method of claim 54, wherein said nucleotides code for a
picornavirus 2C-like protein, a 3C-like protease, an RNA-dependent RNA
15 polymerase or any combination thereof.

80. The probe produced by the method of claim 79.

81. The method of claim 54, wherein said nucleotide codes for a capsid protein.

82. The probe produced by the method of claim 81.

20 83. The method of claim 54, wherein said subgenomic region includes nucleotides 5337 through 7573.

84. The probe produced by the method of claim 83.

85. The method of claim 54, wherein said subgenomic region includes nucleotides 5346 through 6935.

86. The probe produced by the method of claim 85.

87. The method of claim 54, wherein said subgenomic region
5 includes nucleotides 6938 through 7573.

88. The probe produced by the method of claim 87.

89. A method of making a probe to detect Norwalk-related viruses, comprising the steps of:

10 selecting one or more nucleotide sequences from the group
consisting of GTTGCTGTTGGCATTACA,
TAGTGGCATGGGTATTTC, ATTATAGTTTCTTGCATA,
AGCAAAGTCATACATGAAAT, and ACTCACCCAAATCCTCCA;
producing said nucleotide sequence by chemical methods or
in an expression system.

15 90. The probe produced by the method of claim 89.

91. A kit for detecting an immune response to Norwalk virus,
comprising:

a container including a protein encoded by the Norwalk virus
genome shown in Table 2 or fragments or derivatives thereof.

20 92. The kit of claim 91, wherein said protein is selected from the
group consisting of the protein encoded by nucleotides 1 through 7753, the
protein encoded by nucleotides 146 through 5359, the protein encoded by
nucleotides 5337 through 7573, the protein encoded by nucleotides 5346
through 6935, the protein encoded by nucleotides 6938 through 7573 and
25 any combination thereof.

93. A kit for detecting an immune response to a Norwalk-related virus, comprising:

a container including a protein encoded by the genome for said Norwalk-related virus.

5 94. A method of detecting an immune response to Norwalk virus, comprising the steps of:

collecting a serum sample from an individual suspected of having been exposed to Norwalk virus;

10 selecting a protein encoded by the Norwalk virus genome shown in Table 2 or fragments or derivatives thereof;

adding said selected protein to said serum in a diagnostic assay under conditions allowing said selected protein and the serum to react; and

15 measuring the amount of reaction of said serum and said selected protein.

95. The method of claim 94, wherein said diagnostic assay is selected from the group consisting of enzyme-linked immunosorbent assays, radioimmunoassays and immunoblots.

20 96. The method of claim 94, wherein said selected protein is a capsid protein.

97. The method of claim 94, wherein said selected protein has the intrinsic property of being able to form particle(s).

98. The method of claim 94, wherein said selected protein is selected from the group consisting of the protein encoded by nucleotides
25 1 through 7753, the protein encoded by nucleotides 146 through 5359, the protein encoded by nucleotides 5337 through 7573, the protein encoded by

nucleotides 5346 through 6935, the protein encoded by nucleotides 6938 through 7573 and any combination thereof.

99. A diagnostic assay to detect an immune response to Norwalk virus, comprising:

- 5 selecting a protein encoded in Norwalk virus genome shown in Table 2 or fragments or derivatives thereof;
 using said protein as an antigen;
 adding post-infection serum from a Norwalk infected individual under conditions allowing said serum to react with said
10 antigen; and
 measuring the amount of reaction of said serum and said antigen.

100. The method of claim 99, wherein said protein is a capsid protein.

- 15 101. The method of claim 99, wherein said protein has the intrinsic property of being able to form particle(s).

102. The method of claim 99, selected from the group consisting of the protein encoded by nucleotides 1 through 7753, the protein encoded by nucleotides 146 through 5359, the protein encoded by nucleotides 5337
20 through 7573, the protein encoded by nucleotides 5346 through 6935, the protein encoded by nucleotides 6938 through 7573 and any combination thereof.

103. A kit for detecting Norwalk viruses and Norwalk-related viruses, comprising:

- 25 a container including at least one antiserum made from a protein encoded by the Norwalk virus genome shown in Table 2 or from a fragment or derivative of said genome.

104. The kit of claim 102, wherein said protein is selected from the group consisting of the protein encoded by nucleotides 1 through 7753, the protein encoded by nucleotides 146 through 5359, the protein encoded by nucleotides 5337 through 7573, the protein encoded by nucleotides
5 5346 through 6935, the protein encoded by nucleotides 6938 through 7573 and any combination thereof.

105. A method of producing antibodies to Norwalk and Norwalk-related viruses, comprising:

immunizing animals with a protein encoded by the Norwalk
10 virus genome shown in Table 2 or fragments or derivatives thereof.

106. The method of claim 105, wherein said protein is selected from the group consisting of the protein encoded by nucleotides 1 through 7753, the protein encoded by nucleotides 146 through 5359, the protein encoded by nucleotides 5337 through 7573, the protein encoded by
15 nucleotides 5346 through 6935, the protein encoded by nucleotides 6938 through 7573 and any combination thereof.

107. A vaccine for Norwalk virus, comprising:

a Norwalk virus antigen encoded by the cDNA sequence of
Norwalk virus shown in Table 2 or fragments or derivatives
20 thereof.

108. The vaccine of claim 107, wherein said antigen is produced using nucleotides 146 through 5359 of the Norwalk virus genome shown in Table 2 or a derivative thereof.

109. The vaccine of claim 107, wherein said antigen is produced
25 using nucleotides 5337 through 7573 of the Norwalk virus genome shown in Table 2 or a derivative thereof.

110. The vaccine of claim 107, wherein said antigen is produced using nucleotides 5346 through 6935 of the Norwalk virus genome shown in Table 2 or a derivative thereof.

111. The vaccine of claim 107, wherein said antigen is produced
5 using nucleotides 6938 through 7573 of the Norwalk virus genome shown in Table 2 or a derivative thereof.

112. The vaccine of claim 107, wherein said antigen has the intrinsic property of being able to form particle(s).

113. A method of immunizing an individual against Norwalk
10 virus, comprising the step of:
orally or parenterally administering an immunologically effective dose(s) of the vaccine of claim 107.

114. A method of immunizing an individual against Norwalk
virus, comprising the steps of:
15 orally and parenterally administering an immunologically effective dose of the vaccine of claim 107.

115. A cDNA sequence of the human calicivirus Sopporo genome shown in Figure 9 and fragments and derivatives thereof, said fragments and derivatives having sufficient size and nucleotide homology to bind a
20 Norwalk or Norwalk-related virus genome.

116. A protein encoded by nucleotides including nucleotides 1 through 551 of the human calicivirus Sopporo genome shown in Figure 9 or fragments or derivatives thereof.

117. A cDNA subclone of the human calicivirus Sopporo genome
25 comprising nucleotides 1 through 149 and fragments and derivatives

thereof, said fragments and derivatives having sufficient size and nucleotide homology to bind a Norwalk or Norwalk-related virus genome.

118. A cDNA subclone of the human calcivirus Sopporo genome comprising nucleotides 113 through 551 and fragments and derivatives
5 thereof, said fragments and derivatives having sufficient size and nucleotide homology to bind a Norwalk or Norwalk-related virus genome.

119. A cDNA sequence of the Day care calcivirus genome shown in Figure 9 and fragments and derivatives thereof, said fragments and derivatives having sufficient size and nucleotide homology to bind a
10 Norwalk or Norwalk-related virus genome.

120. A cDNA sequence of the SRSV/KY/89 genome shown in Figure 12 and fragments and derivatives thereof, said fragments and derivatives having sufficient size and nucleotide homology to bind a Norwalk or Norwalk-related virus genome.

121. A cDNA sequence of the human calcivirus Houston shown in Table 10 and fragments and derivatives thereof, said fragments and derivatives having sufficient size and nucleotide homology to bind a Norwalk or Norwalk-related virus genome.

122. A cDNA subclone of a primate calcivirus comprising the
20 sequence TGGACGGACC TGCTGTTGAA GATCTCTTCA
AANGGCTCGA ACGACCAAAG CACGATCGGT ATTGTGTTGA
CTACGCAAAG TGGGACTCAA CCCANCCACCA AAAGTAACAT
CCAATCAATN GACATC and fragments and derivatives thereof, said
25 fragments and derivatives having sufficient size and nucleotide homology to bind a Norwalk or Norwalk-related virus genome.

123. A cDNA subclone of a primate calcivirus comprising the sequence GTGANATGNN ACATCTTCGA CTCGATGGAC CTATTCACAT

ATGGTGATGA CGGTGTCTAC ATCGTCCCAC CACTATATCA
TCTGTCATGC CCAAGTCTTC ACCAACCTGA AAC and fragments and
derivatives thereof, said fragments and derivatives having sufficient size
and nucleotide homology to bind a Norwalk or Norwalk-related virus
5 genome.

124. A method of detecting an immune response to Norwalk or a
Norwalk related virus, comprising the steps of:

- collecting a serum sample from an individual suspected of
having been exposed to Norwalk or a Norwalk related virus;
- 10 selecting a protein encoded by the genomic sequence of a
Norwalk-related virus or fragments or derivatives thereof, said
fragments and derivatives having sufficient size and nucleotide
homology to bind a Norwalk or Norwalk-related virus genome;
- adding said selected protein to said serum in a diagnostic
15 assay under conditions allowing the selected protein and the serum
to react; and
- measuring the amount of reaction of said serum and said
selected protein.

125. The method of claim 124, wherein said diagnostic assay is
20 selected from the group consisting of enzyme-linked immunosorbent
assays, radioimmunoassays and immunoblots.

126. The method of claim 124, wherein said genomic sequence is
the cDNA sequence of claim 117.

127. The method of claim 124, wherein said genomic sequence is
25 the cDNA sequence of claim 119.

128. The method of claim 124, wherein said genomic sequence is
the cDNA sequence of claim 120.

129. The method of claim 124, wherein said genomic sequence is the cDNA sequence of claim 121.

130. The method of claim 124, wherein said genomic sequence is the cDNA sequence of claim 122.

5 131. The method of claim 124, wherein said genomic sequence is the cDNA sequence of claim 123.

132. A kit for detecting Norwalk viruses and Norwalk-related viruses, comprising:

10 a container including at least one antiserum made from a protein encoded by genomic sequence of a Norwalk-related virus genome or from a fragment or derivative said genomic sequence, said fragments and derivatives having sufficient size and nucleotide homology to bind a Norwalk or Norwalk-related virus genome.

15 133. The kit of claim 132, wherein said genomic sequence is the cDNA sequence of claim 117.

134. The kit of claim 132, wherein said genomic sequence is the cDNA sequence of claim 119.

135. The kit of claim 132, wherein said genomic sequence is the cDNA sequence of claim 120.

20 136. The kit of claim 132, wherein said genomic sequence is the cDNA sequence of claim 121.

137. The kit of claim 132, wherein said genomic sequence is the cDNA sequence of claim 122.

138. The kit of claim 132, wherein said genomic sequence is the cDNA sequence of claim 123.

139. A chimeric protein, comprising:
a protein encoded by a Norwalk virus genome combined with
5 a protein encoded by a genome of a Norwalk-related virus.

140. A method of detecting an immune response to Norwalk virus, comprising the steps of:
collecting a serum sample from an individual suspected of
having been exposed to Norwalk virus;
10 adding said the chimeric protein of claim 139 to said serum
in a diagnostic assay under conditions allowing chimeric protein
and the serum to react; and
measuring the amount of reaction of said serum and said
chimeric protein.

141. A vaccine for Norwalk or Norwalk related viruses,
15 comprising
the chimeric protein of claim 139 used as an antigen.

142. A kit for detecting Norwalk or Norwalk-related related
viruses, comprising:
20 a container including at least one antiserum made from the
chimeric protein of claim 139.

Figure 1a

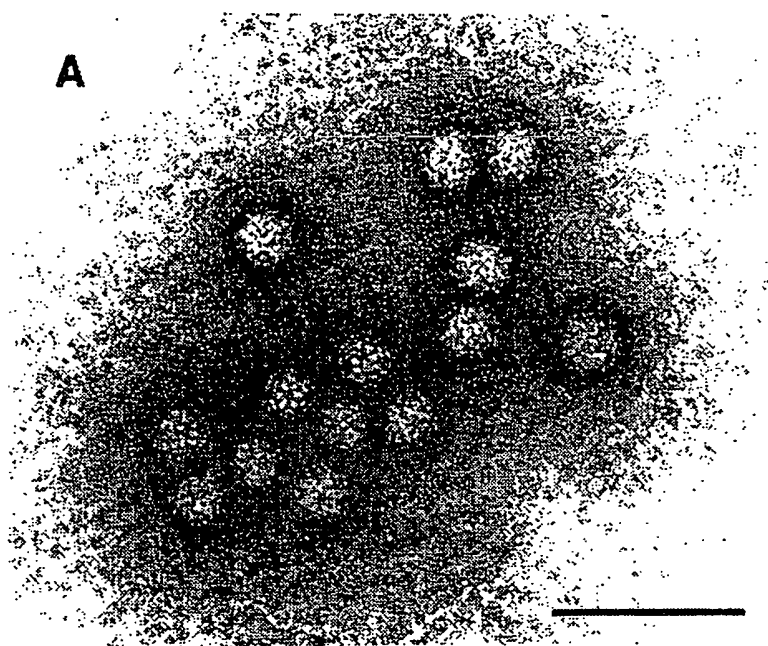
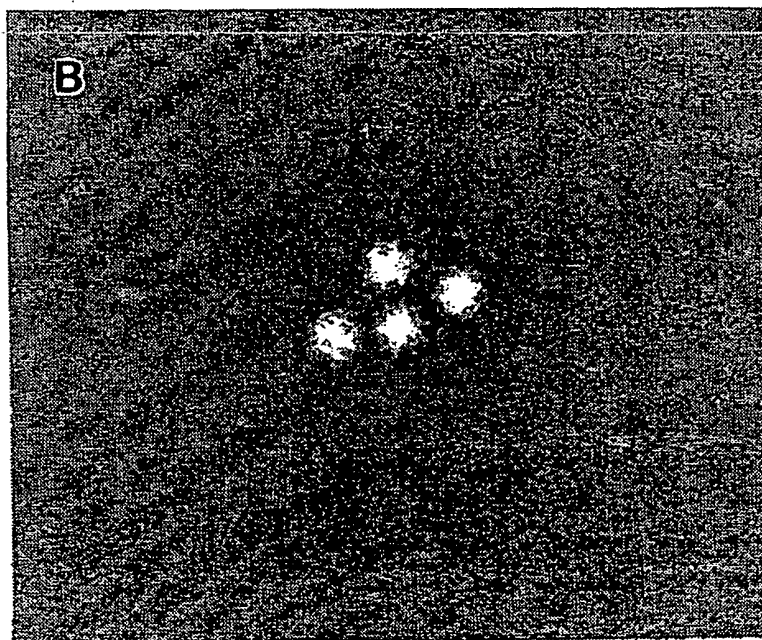
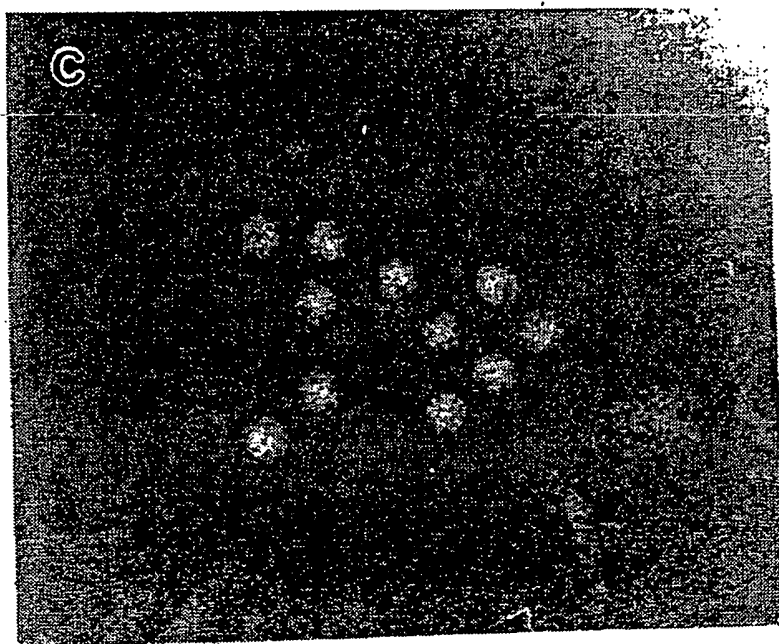


Figure 1b



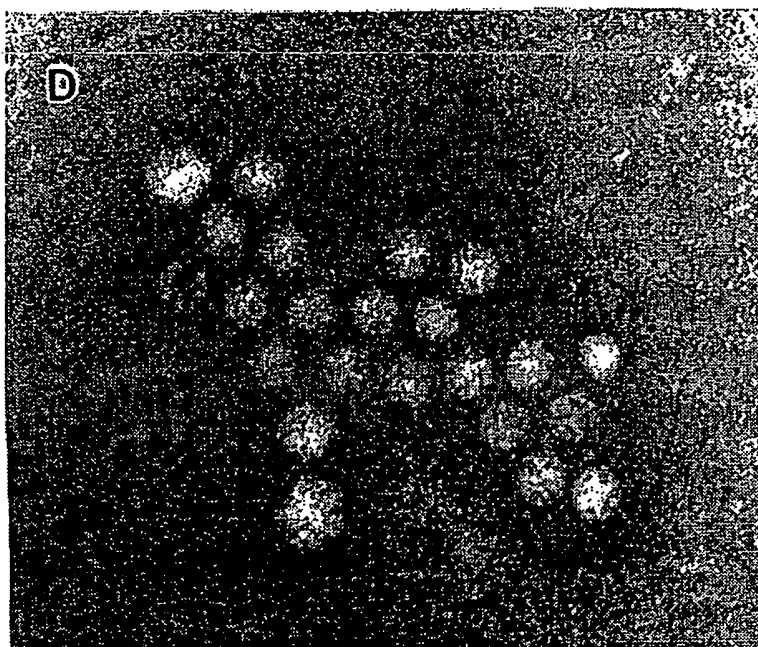
SUBSTITUTE SHEET

Figure 1c



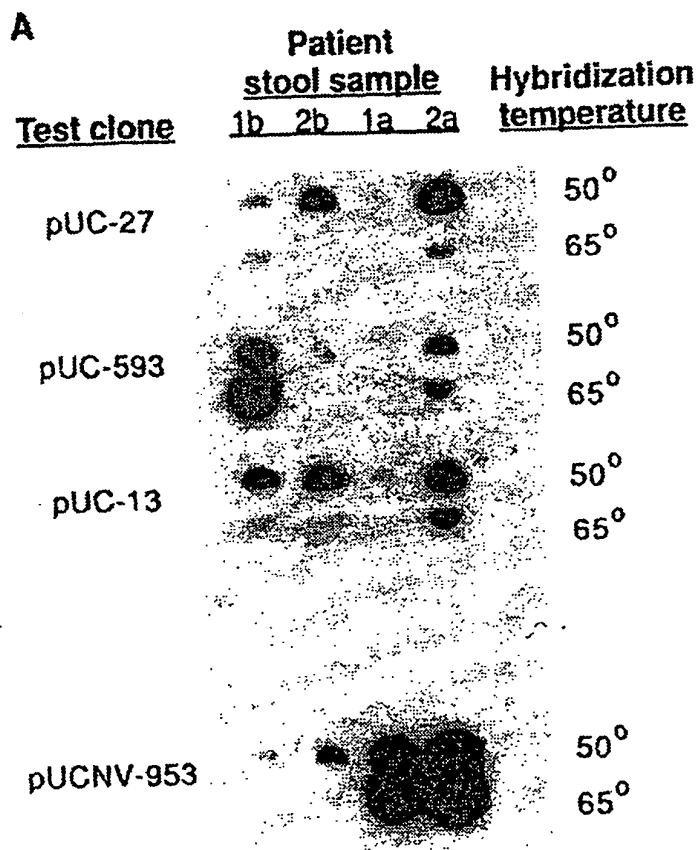
SUBSTITUTE SHEET

Figure 1d



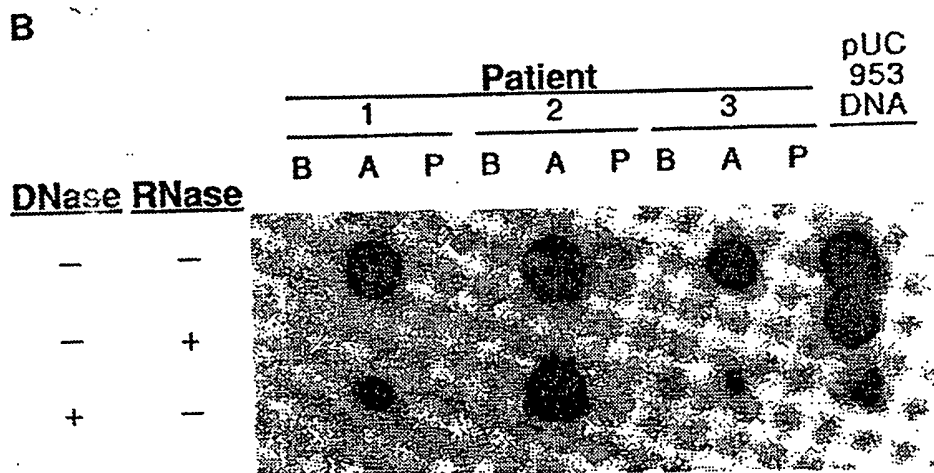
SUBSTITUTE SHEET

Figur 2a



SUBSTITUTE SHEET

Figure 2b



SUBSTITUTE SHEET

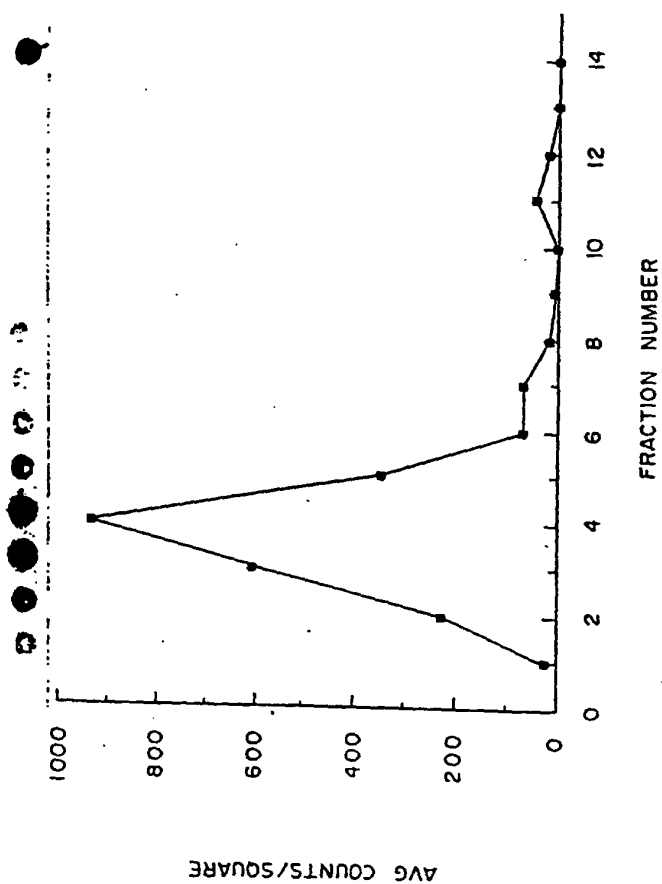


FIG. 3

SUBSTITUTE SHEET

8/39

G TGC TCT GGG AGC GGG CAT ACA GGT TGG TGG CGA CAG GCC CTC CAA
 cys ser gly ser gly his thr gly trp trp arg gln ala leu gln

AGC CAA AGG TAT CAA CAA AAT TTG CAA CTG CAA GAA AAT TCT TTT
 ser gln arg tyr gln gln asn leu gln leu gln glu asn ser phe

AAA CAT GAC AGG GAA ATG ATT GGG TAT CAG GTT GAA GCT TCA AAT
 lys his asp arg glu met ile gly tyr gln val glu ala ser asn

CAA TTA TTG GCT AAA AAT TTG GCA ACT AGA TAT TCA CTC CTC CGT
 gln leu leu ala lys asn leu ala thr arg tyr ser leu leu arg

GCT GGG GGT TTG ACC AGT GCT GAT GCA GCA AGA TCT GTG GCA GGA
 ala gly gly leu thr ser ala asp ala ala arg ser val ala gly

GCT CCA GTC ACC CGC ATT GTA GAT TGG AAT GGC GTG AGA GTG TCT
 ala pro val thr arg ile val asp trp asn gly val arg val ser

GCT CCC GAG TCC TCT GCT ACC ACA TTG AGA TCC GGT GGC TTC ATG
 ala pro glu ser ser ala thr thr leu arg ser gly gly phe met

TCA GTT CCC ATA CCA TTT GCC TCT AAG CAA AAA CAG GTT CAA TCA
 ser val pro ile pro phe ala ser lys gln lys gln val gln ser

TCT GGT ATT AGT AAT CCA AAT TAT TCC CCT TCA TCC ATT TCT CGA
 ser gly ile ser asn pro asn tyr ser pro ser ser ile ser arg

ACC ACT AGT TGG GTC GAG TCA CAA AAC TCA TCG AGA TTT GGA AAT
 thr thr ser trp val glu ser gln asn ser ser arg phe gly asn

CTT TCT CCA TAC CAC GCG GAG GCT CTC AAT ACA GTG TGG TTG ACT
 leu ser pro tyr his ala glu ala leu asn thr val trp leu thr

CCA CCC GGT TCA ACC
 pro pro gly ser thr

FIG 4

SUBSTITUTE SHEET

9/39

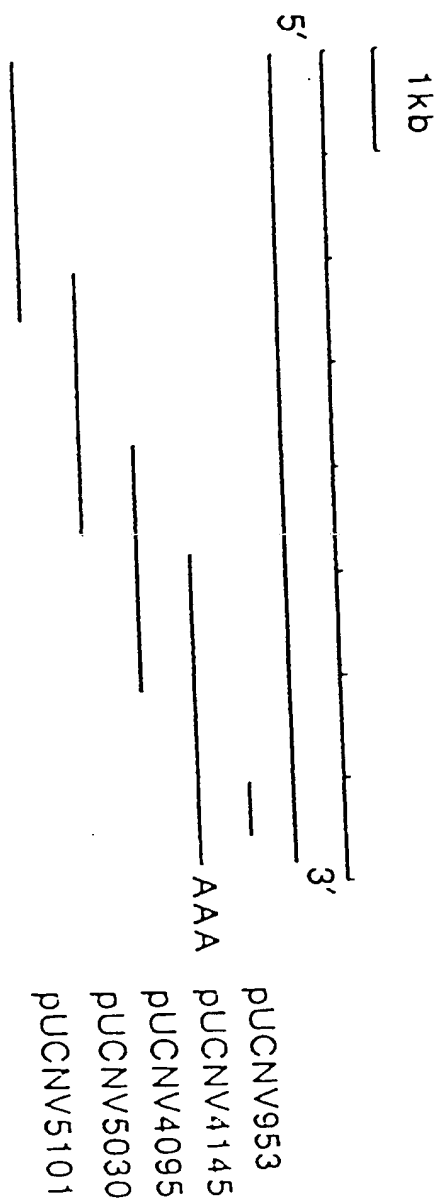


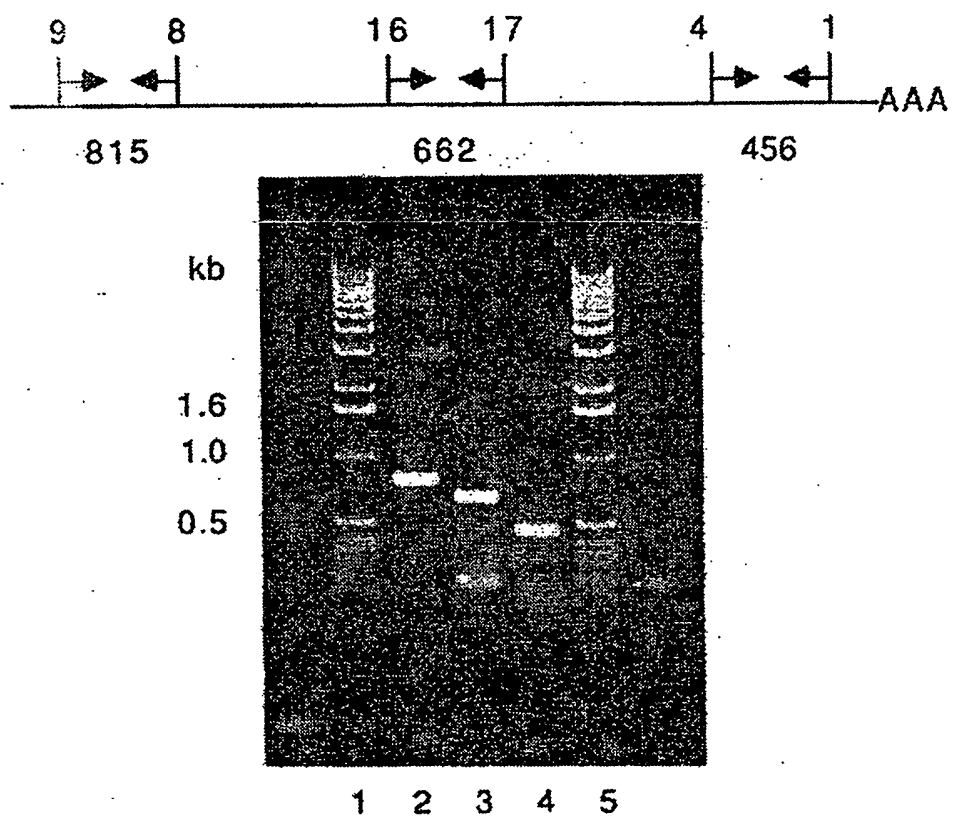
FIG 5

10/39

[illegible]

॥६॥

Figure 7



SUBSTITUTE SHEET

12/39

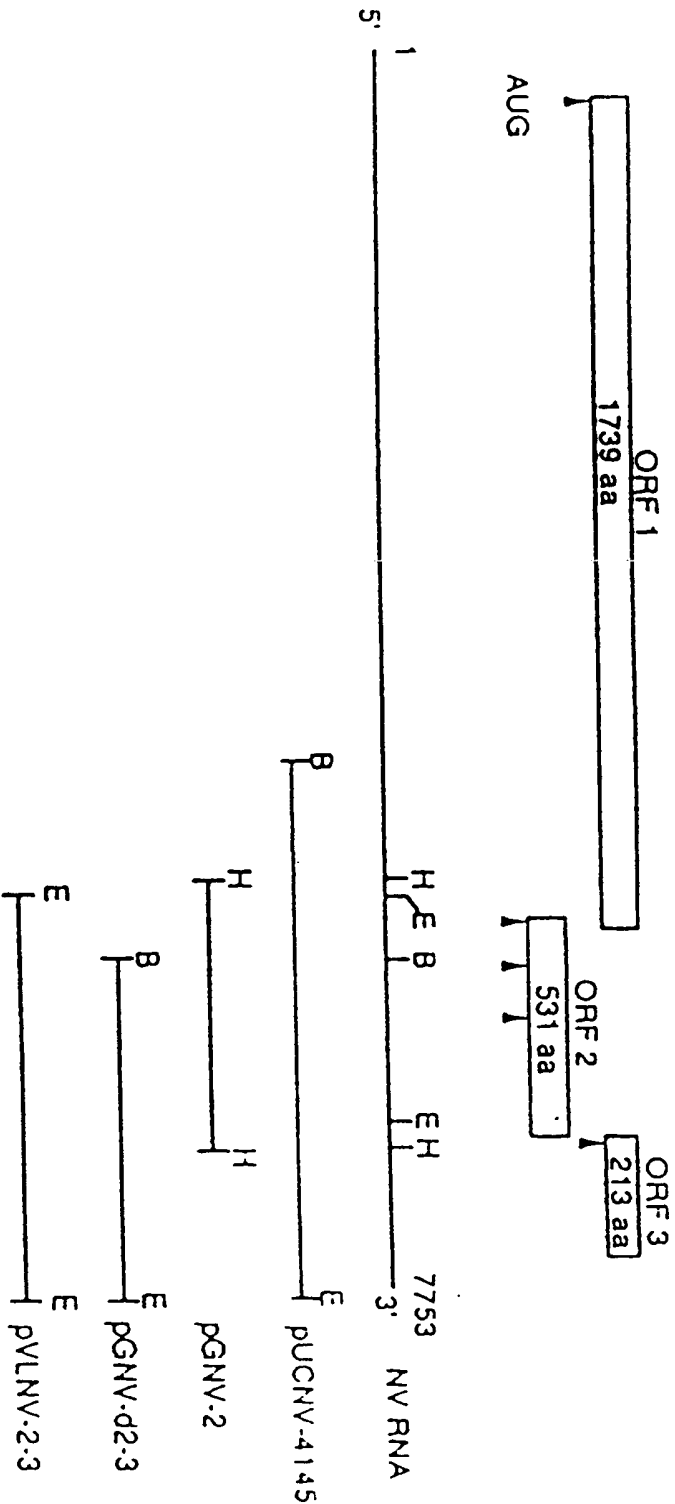


FIG 8

Figure 9(1)

TGT GAT GCT GCC ACC ACG CTT ATA GCC ACC GCG GCT TTT AAG GCC⁴⁵
 C D₁ A A T T L I A T A A F K A
 ←.....a23s2m31.....

GTG GCT ACN AGG CTA CAG GTG GTG ACA CCA ATG ACA CCA GTT GCT⁹⁰
 A V X R L Q V V T P M T P V A

13/39
 GTT GGC ATT AAC ATG GAC TCT GTT CAG ATG CAA GTG ATG AAT GAC¹³⁵
 V G I N M D S V Q M Q V M N D

.....primer new 36.....
 ←.....c-29_4-gcl.....

... ..ACCCCCC¹⁸⁰
 TCT TTA AAG GGG GGT GTT CTT TAC TGT TTG GAT TAT TCC AAA TGG
 S L K G G V L Y C L D Y S K W

SUBSTITUTE SHEET

Figure 9(2)

...CT G... ..AAN N... ..
 GAT TCC ACA CAA AAC CCT GCA GTG ACA GCA GCC TCC CTG GCA ATA
 D S T Q N P A V T A A S L A I
 225 Day care
 HuCV Sapporo
 Sapporo a.a.

... ..ACTCTCAI
 TTG GAG AGA TTT GCT GAG CCC CAT CCA ATT GTG TCT TGT GCC ATT
 L E R F A E P H P I V S C A I
 270 Day care
 HuCV Sapporo
 Sapporo a.a.

... ..G NN... ..N N... ..TTCTCG
 GAG GCT CTT TCC TCC CCT GCA GAG GGC TAT GTC AAT GAT ATC AAA
 E A L S S P A E G Y V N D I K
 315 Day care
 HuCV Sapporo
 Sapporo a.a.

... ..TGGGGGCCC
 TTT GTG ACA CGC GGC GGC GGC CTA CCA TCT GGG ATG CCA TTT ACA TCT
 F V T R G G L P S G M P F T S
 360 Day care
 HuCV Sapporo
 Sapporo a.a.

Figure 9(3)

...T ...C ...C ...N ...C ...A ...C ...
 GTC GTC AAT TCT ATC AAC CAT ATG ATA TAC GTG GCG GCA GCC ATC
 V V N S I N H M I Y V A A A I
 405 Day care
 HuCV Sapporo
 Sapporo a.a.

...CAG GCA TAC GAA AGC CAC AAT GTC CCA TAT ACT GGA AAC GTC
 L G A Y E S H N V P Y T G N V
 450 Day care
 HuCV Sapporo
 Sapporo a.a.

15/39

...CAA GTG GAG ACC GTT CAC ACG TAT GGT GAT GAT TGC ATG TAC
 F Q V E T V H T Y G D D C M Y
 495 Day care
 HuCV Sapporo
 Sapporo a.a.

SUBSTITUTE SHEET

Figure 9(4)

540
 AGC GTG TGC CCT GCC ACT GCA TCA ATT TTC CAC ACT GTG CTT GCA Day care
 S V C P A T A S I F H T V L A HuCV Sapporo
 Sapporo a.a.

551
 AAC CTA ACG TC HucV Sapporo
 N L T Sapporo a.a.

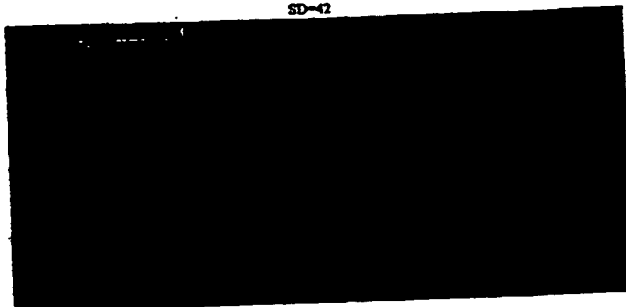
16/39

→

Figure 10

Sapporo

SD-42



Day Care

Houston

SD-3

SD-3

Primate

SD-17



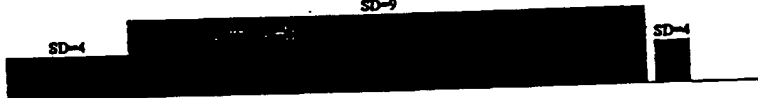
Feline F9

SD-9

SD-4

SD-4

Rabbit



SD-2

Norwalk



4394

4777

SUBSTITUTE SHEET

19/39

Figure 12 (1)

4494 -	CAATAGAAGA	TGGCCCCCTC	ATCTATGCTG	AGCATGCTAA	ATATAAGAAT	CATTTTGATG	Norwalk virus
							SRSV/KY/89
1 -	CAATAGAGGA	TGGCCCCCTTA	ATTATGCTG	AACATGCCAA	GTACAAAAT	CATTTTGATG	
4554 -	CAGATTATAC	AGCATGGGAC	TCAACACAAA	ATAGACAAAT	TATGACAGAA	TCCTTCTCCA	Norwalk virus
							SRSV/KY/89
61 -	CAGATTACAC	AGCATGGGAC	TCTACACAAA	ATAGACAAAT	TATGACAGAA	TCCTTCTCCA	
4614 -	TTATGTGCGG	CCTTACGGCC	TCACCAGAAT	TGGCCGAGGT	TGTGGCCCCA	GATTTGCTAG	Norwalk virus
							SRSV/KY/89
121 -	TCATGTACG	CCTTACGGCC	TCTCCAGAAC	TAGCTGAGGT	TGTAGCCCCG	GACTTACTAG	
4674 -	CACCATCTGA	GATGGATGTA	GGTGATTATG	TCATCAGGGT	CAAAGAGGGG	CTGCCATCTG	Norwalk virus
							SRSV/KY/89
181 -	CACCATCCGA	GATGGATGTG	GCGACTATG	TTATAAGSGT	CAAAGAAGGC	CTACCATCAG	
4734 -	GATTCOCATG	TACTTCCCAG	GTGAACAGCA	TAAATCACTG	GATAATTACT	CTCTGTGCAC	Norwalk virus
							SRSV/KY/89
241 -	GATTTCCTCG	CACCTTCTCA	GTGAATAGCA	TAAATCACTG	GATAATCACC	CTTTGTGCAT	
4794 -	TGCTGAGGC	CACGTGGTTTA	TCACCTGATG	TGGTGCAATC	CATGTCATAT	TTCTCATTTT	Norwalk virus
							SRSV/KY/89
301 -	TGCTGAGGC	TACTGGCTTA	TCACCTGATG	TGGTACAGTC	CATGTCATAC	TTCTCATTCT	
							*
4854 -	ATGGTGATGA	TGAGATTGTG	TCAAACAGCA	TAGATTTTGA	CCCAGCCCCG	CTCACTCAAA	Norwalk virus
							SRSV/KY/89
361 -	ACGGTGATGA	TGAGATCGTA	TCAAACAGCA	TAGACTTTGA	CCCAACTCGC	CTCACCCCAA	
	*****	****					

SUBSTITUTE SHEET

20/39

Figure 12(2)

4914	-	TTCTCAAGGA	ATATGGCCTC	AAACCAACAA	GGCCTGACAA	AACAGAAGGA	CCAATACAAG	Norwalk virus
421	-	TTCTCAAGGA	ATACGGCCTC	AAGCCAACAA	GGCCAGACAA	AACAGAAGGA	CCAATACAGG	SRSV/KY/89
		+++++	+++++	+++++	+++++	+++++	+++++	
4974	-	TGAGGAAAAA	TGTGGATGGA	CTGGTCTTCT	TGCGGCGCAC	CATTTCCTCGT	GATCGGCGCAG	Norwalk virus
481	-	TGAGGAAGAA	TGTGGATGGG	CTAGTTTTTC	TGCGGCGCAC	CATCTCCCGG	GACGCAGCAG	SRSV/KY/89
5034	-	GGTTCCAAGG	CAGGTTAGAT	AGGGCTTCCA	TTGAAACGCCA	AATCTTCTGG	ACCCGCGGGC	Norwalk virus
541	-	GGTTCCNAGG	TAGACTGGAT	AGAGCCTCAA	TTGAAACGTCA	AATTITCTGG	ACCCGCGGGC	SRSV/KY/89
5094	-	CCAATCATTC	AGATCCATCA	GAGACTCTAG	TGCCACACAC	TCAAAGAAAA	ATACAGTTGA	Norwalk virus
601	-	CCAACCATTG	AGACCCATCA	GAGACTCTGG	TACCACACAC	CCAAAGGAAA	GTCCAGCTGA	SRSV/KY/89
5154	-	TTTCACTTCT	AGGGGAAGCT	TCACCTCCATG	GTGAGAAAAT	TTACAGAAAG	ATTTCCAGCA	Norwalk virus
661	-	TCTCACTTCT	AGGAGAAGCC	TCACCTCCACG	GGGAAAAAAT	TTACAGGAAA	ATATCTAGCA	SRSV/KY/89
5214	-	AGGTCATACA	TGAATCAAG	ACTGGTGGAT	TGGAAATGTA	TGTCCCAGGA	TGGCAGGCCA	Norwalk virus
721	-	AAGTCATACA	TGAATTAAG	ACTGGTGGGC	TGGAGATGTA	TGTCCCAGGG	TGGCAGGCCA	SRSV/KY/89
5274	-	TGTTCCGCTG	GATGGCCTTC	CATGACCTCG	GATTGTGGAC	AGGAGATCGC	GATCTTCTGC	Norwalk virus
781	-	TGTTCCGCTG	GATGGCCTTC	CATGACCTCG	GATTGTGGAC	AGGAGATCGC	AATCTCCTGC	SRSV/KY/89

SUBSTITUTE SHEET

21/39

Figure 12 (3)

5334 -	CCGAATTTCGT	AAATGATGAT	GGCGTCTAAG	GACGCTACAT	CAAGCGTGGA	TGGCGCTAGT	Norwalk virus
841 -	CCGAATTTCGT	AAATGATGAT	GGCGTCTAAG	GACGCTACAT	CAAGCGTGGA	TGGCGCCAGT	SRSV/KY/89
	xxx						
5394 -	GGCGCTGGTC	AGTTGGTACC	GGAGGTTAAT	GCTTCTGACC	CTCTTGCAAT	GGATCCTGTA	Norwalk virus
901 -	GGTCGGTTC	AGTTGGTACC	GGAGGTTAAT	GCTTCTGACC	CTCTTGCAAT	GGATCCTGTC	SRSV/KY/89
5454 -	GCAGGTTCTT	CGACAGCAGT	CGCGACTGCT	GGACAAGTTA	ATCCTATTGA	TCCCTGGATA	Norwalk virus
961 -	GCAGGTTCTT	CAACAGCAGT	TGCAACCGCT	GGACAAGTTA	ACCCATTATGA	CCCTTGGATA	SRSV/KY/89
5514 -	ATTAATAATT	TTGTGCAAGC	CCCCCAAGGT	GAATTTACTA	TTTCCCCAAA	TAATACCCCC	Norwalk virus
1021 -	ATCAATAACT	TTGTGCAGGC	TCCCCAAGGT	GAATTTACTA	TTTCTCCAAA	TAATACCCCC	SRSV/KY/89
5574 -	GGTGATGTTT	TGTTTGATTT	GAGTTTGGGT	CCCCATCTTA	ATCCTTTCTT	GCTCCATCTA	Norwalk virus
1081 -	GGTGATGTTT	TGTTTGATTT	GAGTCTAGGC	CCTCATCTTA	ATCCCTTCTT	GTTACATTGT	SRSV/KY/89
5634 -	TCACAAATGT	ATAATGGTTG	GTTTGGTAAC	ATGAGAGTCA	GGATTATGCT	AGCTGGTAAT	Norwalk virus
1141 -	TCACAAATGT	ATAATGGCTG	GGTTGGCAAC	ATGAGAGTTA	GGATTATGCT	GGCTGGTAAT	SRSV/KY/89
5694 -	GCTTTTACTG	CGGGGAAGAT	AATAGTTTCC	TGCATACCCC	CTGGTTTTTG	TTCACATAAT	Norwalk virus
1201 -	GCAATTTACTG	CAGGCAAAAT	TATAGTTTCT	TGCATACCTC	CTGGCTTTTG	CTCCCAACAA	SRSV/KY/89
5754 -	CTTACTATAG	CACAAGCAAC	TCTCTTTCCA	CATGTGATTG	CTGATGTTAG	GACTCTAGAC	Norwalk virus
1261 -	CTTACTATAG	CACAAGCAAC	TCTCTTCCCG	CATGTGATTG	CTGATGTTAG	GACTTTAGAC	SRSV/KY/89

SUBSTITUTE SHEET

22/39

Figure 12 (4)

5814 -	CCCATTTGAGG	TGCCCTTTTGA	AGATGTTAGG	AATGTTCTCT	TTTATAATAA	TGATAGAAAT	Norwalk virus
							SRSV/KY/89
1321 -	CCATTTGAAG	TACCCCTTGA	AGATGTAAGG	AATGTTCTCT	TTTATAATAA	TGATAGAAAT	Norwalk virus
							SRSV/KY/89
5874 -	CAACAAACCA	TGCGCCTTGT	GTGCATGCTG	TACACCCCCC	TCCGCACTGG	TGGTGGTACT	Norwalk virus
							SRSV/KY/89
1381 -	CAACAAACTA	TGCGCCTTGT	GTGCATGCTT	TATACCCCCC	TCAGCACTGG	TGGCGGTACA	Norwalk virus
							SRSV/KY/89
5934 -	GGTGATTCTT	TTGTAGTTGC	AGGGCGAGTT	ATGACTTGCC	CCAGTCCCTG	TTTTAATTTC	Norwalk virus
							SRSV/KY/89
1441 -	GGTGATTCTT	TTGTGGTTGC	AGGGCGAGTC	ATGACTTGTC	CTAGCCCCCG	CTTTAATTTC	Norwalk virus
							SRSV/KY/89
5994 -	TTGTTTTTAG	TCCCTCCTAC	GGTGGAGCAG	AAAACCAGGC	CCTTCACACT	CCCAAACTCTG	Norwalk virus
							SRSV/KY/89
1501 -	TTGTTCTTGG	TTCCCTCCAC	AGTGAACAG	AAGACTAGGC	CTTTCACCCCT	CCCAAAATTA	Norwalk virus
							SRSV/KY/89
6054 -	CCATTGAGTT	CTCTGTCTAA	CTCAGGTGCC	CCTCTCCCAA	TCAGTAGTAT	CGGCATTTC	Norwalk virus
							SRSV/KY/89
1561 -	CCGCTGAGTT	CTTTGTCTAA	TTTACGTGCT	CCTCTTCCAA	TTAGTGGCAT	GGGTATTCT	Norwalk virus
							SRSV/KY/89
6114 -	CCAGACAATG	TCCAGAGTGT	GCAGTTCCAA	AATGGTCGGT	GTACTCTGGA	TGGCCGCCCTG	Norwalk virus
							SRSV/KY/89
1621 -	CCAGATAATG	TTTACAGAGTGT	GCAGTTCCAA	AATGGCCGAT	GTACCTTAGA	TGGACGTCTT	Norwalk virus
							SRSV/KY/89
6174 -	GTTGGCACCA	CCCCAGTTTC	ATTGTCACAT	GTGCCCCAAGA	TAAGAGGGAC	CTCCAATGSC	Norwalk virus
							SRSV/KY/89
1681 -	GTTGGCACCA	CCCCAGTTTC	CCTCTCCCAT	GTGCTAAGA	TAAGGGGTAC	TTCTAATGGT	Norwalk virus
							SRSV/KY/89
6234 -	ACTGTAATCA	ACCTTACTGA	ATTGGATGGC	ACACCCCTTC	ACCCTTTTGA	GGGCCCTGCC	Norwalk virus
							SRSV/KY/89
1741 -	ACAGTAATCA	ATCTCACCGA	ATTGGATGGC	ACCCCTTTC	ACCCTTTTGA	AGGCCCTGCC	Norwalk virus
							SRSV/KY/89

UBSTITUTE SHEET

23/39

Figure 12 (5)

6294 -	CCCATTTGGGT	TTCCAGACCT	CGGTGGTTGT	GATTGGCATA	TCAATATGAC	ACAGTTTGGC	Norwalk virus
1801 -	CCTATTGGTT	TTCCAGATCT	TGGTGGCTGT	GATTGGCATA	TTAATATGAC	ACAATTTGGA	SRSV/KY/89
6354 -	CATTCTAGCC	AGACCCAGTA	TGATGTAGAC	ACCACCCCTG	ACACTTTTGT	CCCCCATCTT	Norwalk virus
1861 -	CATTCCAGTC	AGACTCAGTA	TGATGTAGAC	ACCACCCCTG	ACACCTCCGT	CCCTCACCTA	SRSV/KY/89
6414 -	GGTTCAATTC	AGGCAATGG	CATTGGCAGT	GGTAATTATG	TTGGTGTCT	TAGCTGGATT	Norwalk virus
1921 -	GGTTCAATCC	AGCGGAATGG	CATTGGTAGT	GGCAACTATA	TTGGTGTCT	TAGCTGGGTC	SRSV/KY/89
6474 -	TCCCCCCCCAT	CACACCCGTC	TGGCTCCCAA	GTTGACCTTT	GGAAGATCCC	CAATTATGGG	Norwalk virus
1981 -	TCCCCCCCCAT	CACATCCATC	TGGCTCTCAA	GTTGATCTCT	GGAAGATCCC	CAACTATGGG	SRSV/KY/89
6534 -	TCAAGTATTA	CGGAGGCAAC	ACATCTAGCC	CCTTCTGTAT	ACCCCCCTGG	TTTCGGGAGAG	Norwalk virus
2041 -	TCTAGTATCA	CAGAGGCAAC	CCATCTAGCT	CCCTCTGTCT	ATTCTCCTGG	CTTTGGGAGAG	SRSV/KY/89
6594 -	GTATTGGTCT	TTTTCATGTC	AAAAATGCCA	GGTCCTGGTG	CTTATAATTT	GCCCTGTCTA	Norwalk virus
2101 -	GTGCTAGTCT	TTTTCATGTC	AAAGATACCA	GGTCCTGGTG	GTGATAGTCT	GCCCTGTCTA	SRSV/KY/89
6654 -	TTACCACAAG	AGTACATTTC	ACATCTTGCT	AGTGAACAAG	CCCCTACTGT	AGGTGAGGCT	Norwalk virus
2161 -	CTGCCACAAG	GATATATCTC	ACACCTTGCA	AGTGAACAAG	CCCCAACTGT	TGGTGAGGGT	SRSV/KY/89
6714 -	GCCCTGCTCC	ACTATGTTGA	CCCTGATACC	GGTCGGAATC	TTGGGGAATT	CAAAGCATAC	Norwalk virus
2221 -	CCCCTGCTCC	ACTATGTTGA	CCCTGACACG	GACCGGAATC	TTGGGGAATT	TAAGGCTTAC	SRSV/KY/89

SUBSTITUTE SHEET

Figure 12 (6)

6774 -	CCTGATGGTT	TCCTCACTTG	TGTCCCAAT	GGGGTAGCT	CGGGTCCACA	ACAGTGCCG	Norwalk virus
2281 -	CCTGATGGTT	TCCTAACCTG	TGTCCCTAAT	GGGGCCAGCT	CGGGCCCAACA	ACAACTACCA	SRSV/KY/89
6834 -	ATCAATGGGG	TCTTTGTCTT	TGTTTCATGG	GTGTCCAGAT	TTTATCAATT	AAAGCCTGTG	Norwalk virus
2341 -	ATCAATGGAG	TCTTTGTCTT	TGTTTCATGG	GTGTCCAGAT	TTTATCAGTT	AAAGCCTGTG	SRSV/KY/89
6894 -	GGAAC TGCCA	GCTGGGCAAG	AGGTAGGCTT	GGTCTGGCC	GATAATGGCC	CAAGCCATAA	Norwalk virus
2401 -	GGAAC TGCCA	GTACGGCAAG	AGGTAGGCTT	GGTTGGGCC	GATAATGGCT	CAGGCTATAA	SRSV/KY/89
6954 -	TTGGTGCAAT	TGCTGCTTCC	ACAGCAGGTA	GTGCTCTGGG	AGCGGGCATA	CAGGTT	Norwalk virus
2461 -	TTGGTGCAAT	TGCCGCCTCT	ACAGCAGGTA	GTGCTTTAGG	GGCAGGTATA	CAGGTT	SRSV/KY/89

###

* Location of YGDD motif
+ Location of primer 35
x Beginning of capsid protein
Beginning of third ORF

Parameters for homology search:

Number of diagonals searched: 25
Pre-processing ktuple value: 2
DD algorithm used to find homologies on diagonals (kmatch = 1)
Scoring matrix used: Unit
Threshold SD score for saving homology domains: 3.0
Minimum length for homology domains: 15

SUBSTITUTE SHEET

25/39

Figure 12 (7)

S score: Similarity score of homology domain
 SD Score: (Similarity score - expected score) / std. dev.
 Expected similarity score: 4.99
 Std. Dev. of similarity scores: 0.43
 Threshold SD score for display of homology domains: 3.0
 Equivalent threshold S_score: 6.3

Comparison:

C1 (1f): |>u 1>+++++ Norwalk virus 58gelcorr8 (7644 bases)++++>u 7644>|
 C2 (1f): |>u 1>+++++ SRSV/KY/89-p36-p3 (2516 bases)++++>u 2516>|

SD_Score	S_Score	%similar	length	limits_seq_1	limits_seq_2
155.8	72.0	87.2	2516	4494 -> 7009	1 -> 2516

SUBSTITUTE SHEET

Figure 13a

1	-	MMMASKDATS	SVDGASASVQ	LVPEVNASDP	LAMPVAGSS	TAVATAGQVN	PIDPWIINNF	SRSV/KY/89
1	-	MMMASKDATS	SVDGASGAGQ	LVPEVNASDP	LAMPVAGSS	TAVATAGQVN	PIDPWIINNF	Norwalk virus
61	-	VQAPQGEFTI	SPNNTPGDVL	FDSLGLPHLN	PFLHLISQMY	NGWVGMRVR	IMLAGNAFTA	SRSV/KY/89
61	-	VQAPQGEFTI	SPNNTPGDVL	FDSLGLPHLN	PFLHLISQMY	NGWVGMRVR	IMLAGNAFTA	Norwalk virus
121	-	GKIIVSCIPP	GFGSQQLTIA	QATLFPHVIA	DVRTLDPIEV	PLEDVRNVLF	HNNDNRNOQTM	SRSV/KY/89
121	-	GKIIVSCIPP	GFGSHNLTIA	QATLFPHVIA	DVRTLDPIEV	PLEDVRNVLF	HNNDNRNOQTM	Norwalk virus
181	-	RLVCMLYTPL	STGGGTGDSF	VVAGRVMTCP	SPDFNLFVLV	PPTVEQKTRP	FTLPNLPSS	SRSV/KY/89
181	-	RLVCMLYTPL	RTGGGTGDSF	VVAGRVMTCP	SPDFNLFVLV	PPTVEQKTRP	FTLPNLPSS	Norwalk virus
241	-	LSNSRAPLPI	SGMGISPDNV	QSVQFQNGRC	TLDGRLVGT	PVSLSHVAKI	RGTSNGTVIN	SRSV/KY/89
241	-	LSNSRAPLPI	SSIGISPDNV	QSVQFQNGRC	TLDGRLVGT	PVSLSHVAKI	RGTSNGTVIN	Norwalk virus
301	-	LTELDTGTPFH	PFEGPAPIGF	PDLGGCDWHI	NMTQFGHSSQ	TQYDVDTTTPD	TSVPHLGSIQ	SRSV/KY/89
301	-	LTELDTGTPFH	PFEGPAPIGF	PDLGGCDWHI	NMTQFGHSSQ	TQYDVDTTTPD	TSVPHLGSIQ	Norwalk virus
361	-	ANGIGSGNVI	GVLSWVSPPS	HPSGSQVDLW	KIPNYGSSIT	EATHLAPSVY	SPGFGEVLVF	SRSV/KY/89
361	-	ANGIGSGNVI	GVLSWISPPS	HPSGSQVDLW	KIPNYGSSIT	EATHLAPSVY	PPGFGEVLVF	Norwalk virus
421	-	FMSKIPGPGG	DSLPCLLPQG	YISHLASEQA	PTVGEGLLH	YVDPDTRNL	GEFKAYPDGF	SRSV/KY/89
421	-	FMSKMPGPGA	YNLPCLLPQE	YISHLASEQA	PTVGEAALLH	YVDPDTRNL	GEFKAYPDGF	Norwalk virus

SUBSTITUTE SHEET

Figure 13a (continued)

SRSV/KY/89
Norwalk virus

481 - LTCVPNGASS GPQQLPINGV FVFVSWSRF YQLKPVGTAS TARGRLGLRR
481 - LTCVPNGASS GPQQLPINGV FVFVSWSRF YQLKPVGTAS SARGRLGLRR

Parameters for homology search:

Number of diagonals searched: 25
Pre-processing ktuple value: 1
DD algorithm used to find homologies on diagonals (kmatch = 1)
Scoring matrix used: Dayhoff
Threshold SD score for saving homology domains: 3.0
Minimum length for homology domains: 15

S_Score: Similarity score of homology domain
SD_Score: (Similarity score - expected score) / std. dev.
Expected similarity score: 3.14
Std. Dev. of similarity scores: 0.87
Threshold SD score for display of homology domains: 3.0
Equivalent threshold S_Score: 5.8

Comparison:

P1 : N>u 1>~~~~ SRSV/KY/89-p36-p3 853 (531 aa)~~~~>u 531>C
P2 : N>u 1>~~~~ Norwalk virus-pep-2 (530 aa)~~~~>u 530>C

SD_Score	S_Score	%similar	length	limits_seq_1	limits_seq_2
51.7	48.1	96.0	530	1 -> 530	1 -> 530

28/39

Figure 13b

1	-	IEDGPLIYAE	HAKYKNHFDA	DYTAWDSTQN	RQIMTESFSI	MSRLTASPEL	AEVVAQDLLA	SRSV/KY/89
								Norwalk virus
1451	-	IEDGPLIYAE	HAKYKNHFDA	DYTAWDSTQN	RQIMTESFSI	MSRLTASPEL	AEVVAQDLLA	
61	-	PSEMDVG DYV	IRVKEGLPSG	FPCTSQVNSI	NHWIITL CAL	SEATGLSPDV	VQSMSYFSFY	SRSV/KY/8
								Norwalk virus
1511	-	PSEMDVG DYV	IRVKEGLPSG	FPCTSQVNSI	NHWIITL CAL	SEATGLSPDV	VQSMSYFSFY	
								+
121	-	GDDEIVSTDI	DFDPTRLTQI	LKEYGLKPTR	PKTEGPIQV	RKNVDGLVFL	RRTISRDAAG	SRSV/KY/8
								Norwalk virus
1571	-	GDDEIVSTDI	DFDPARLTQI	LKEYGLKPTR	PKTEGPIQV	RKNVDGLVFL	RRTISRDAAG	
								++
181	-	FQGR LDRASI	ERQIFWTRGP	NHSDPSETLV	PHTQK VQLI	SLLGEASLHG	EKFYRKISSK	SRSV/KY/8
								Norwalk virus
1631	-	FQGR LDRASI	ERQIFWTRGP	NHSDPSETLV	PHTQK IQLI	SLLGEASLHG	EKFYRKISSK	
					*			
241	-	VIHEIKTGGL	EMVVPGWQAM	FRWRFHDLG	LWTGDRNLLP	EFVND D		SRSV/KY/8
								Norwalk virus
1691	-	VIHEIKTGGL	EMVVPGWQAM	FRWRFHDLG	LWTGDRD LLP	EFVND D		
					*			

+ Location of YGDD motif.

UBSTITUTE SHEET

29/39

Figure 13b (continued)

Parameters for homology search:
 Number of diagonals searched: 25
 Pre-processing ktuple value: 1
 DD algorithm used to find homologies on diagonals (kmatch = 1)
 Scoring matrix used: Dayhoff
 Threshold SD score for saving homology domains: 3.0
 Minimum length for homology domains: 15

S score: Similarity score of homology domain
 SD_Score: (Similarity score - expected score) / std. dev.
 Expected similarity score: 3.14
 Std. Dev. of similarity scores: 0.87
 Threshold SD score for display of homology domains: 3.0
 Equivalent threshold S_Score: 5.8

Comparison:

P1 : N>u 1>~~~~~ SRSV/KY/89-p36-61_3 (289 aa)~~~~~>u 289>C
 P2 : N>u 1>~~~~~ Norwalk virus-pep-1 (1739 aa)~~~~~>u 1739>C

SD_Score	S_Score	%similar	length	limits_seq_1	limits_seq_2
39.4	37.4	99.0	286	1 -> 286	1451 -> 1736

SUBSTITUTE SHEET

Figure 14a

ATGCACTTCA	CAGGTGAATA	GCATCAACCA	CTGGATCCTA	40	SRSV/CDC 6/91
ATGTACCTCA	CAAGTGAACA	GCATCAATCA	CTGGATTTTG	40	SRSV/UT/88
CTGCACATCA	CAGTGGAAAT	CCA-TGCCCA	CTGGCTCCTC	39	SMA/78
CTGCACCTCA	CAGTGGAACT	CCATTGCCCA	CTGGTTGCTT	40	SRSV/Cambridge
TTGCACCTCA	CAGTGGAACT	CCATTGCCCT	CTGGTTGCTT	40	SRSV/CDC 32
ATGTACTTCC	CAGGTGAACA	GCATAAATCA	CTGGATAATT	40	NV/8FIIa/68
CTGCACTTCT	CAAGTAAATA	GCATAAATCA	CTGGATAATC	40	SRSV-3/88
CTGCACTTCT	CAAGTGAATA	GCATAAATCA	CTGGATAATC	40	SRSV/KY89/89
** ** *	** **	** *	**** *		
ACTCTATGTG	CATTGTCTGA	AGTCACTGGC	TTGTCCCCTG	80	SRSV/CDC 6/91
ACCTTGTGGG	GGCTATCAGA	AGTTACTGGT	CTGGCTCCTG	80	SRSV/UT/88
ACACTCTGTG	CACTATCTGA	AGTCACAAAC	CTGGCTCCTG	79	SMA/78
ACTCTGTGTG	CCCTTTCTGA	AGTGACAGGA	CTAGGCCCCG	80	SRSV/Cambridge
ACTCTGTGTG	CCCTTTCTGA	AGTGACAGGA	CTAGGCCCCG	80	SRSV/CDC 32
ACTCTCTGTG	CACTGTCTGA	GGCCACTGGT	TTATCACCTG	80	NV/8FIIa/68
ACCCTTTGTG	CACTGTCTGA	GGCTACTGGC	TTATCACCTG	80	SRSV-3/88
ACCCTTTGTG	CATTGTCTGA	GGCTACTGGC	TTATCACCTG	80	SRSV/KY89/89
** * ** *	* ** *	* **	* ** *		
ATGTGATACA	ATCACAATCT	TATTTCTCAT	TTTATGGT	118	SRSV/CDC 6/91
ATGTAATACA	GTCACAATCT	TACTTTTCAT	TCTATGGT	118	SRSV/UT/88
ACATCATACA	AGCTAACTCC	TTGTTCTCTT	TCTATGGT	117	SMA/78
ACATCATACA	AGCTAATTCC	ATGTACTCTT	TCTATGGT	118	SRSV/Cambridge
ACATCATACA	AGCTAATTCC	ATGTACTCTT	TCTATGGT	118	SRSV/CDC 32
ATGTGGTGCA	ATCCATGTCA	TATTTCTCAT	TTTATGGT	118	NV/8FIIa/68
ATGTGGTGCA	GTCCATGTCA	TACTTCTCAT	TTTACGGT	118	SRSV-3/88
ATGTGGTACA	GTCCATGTCA	TACTTCTCAT	TCTACGGT	118	SRSV/KY89/89
* * * **	* **	* ** *	* ** ** *		

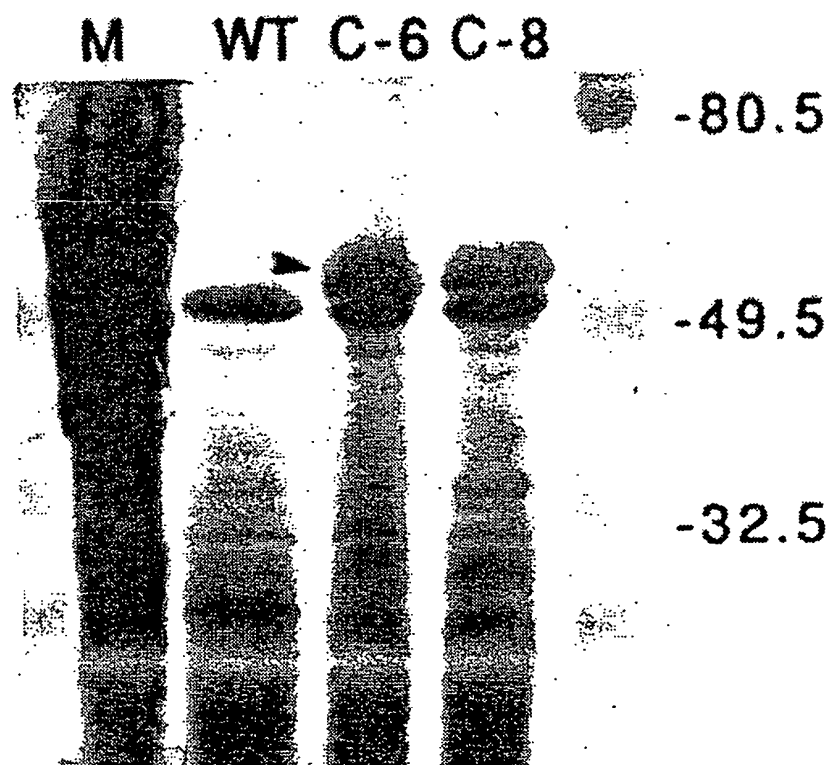
UBSTITUTE SHEET

Figure 14b(3)

ATTCTCAAGGAA	431	NV/8FIIa/68
ATCCTCAAGGAA	431	SRSV-3/88
ATTCTCAAGGAA	431	SRSV/KY89
AAACTCAAAGA-	430	SRSV/Cambridge
AAACTC-----	423	SMA/78
* **		

34/39

Figure 15



SUBSTITUTE SHEET

35/39

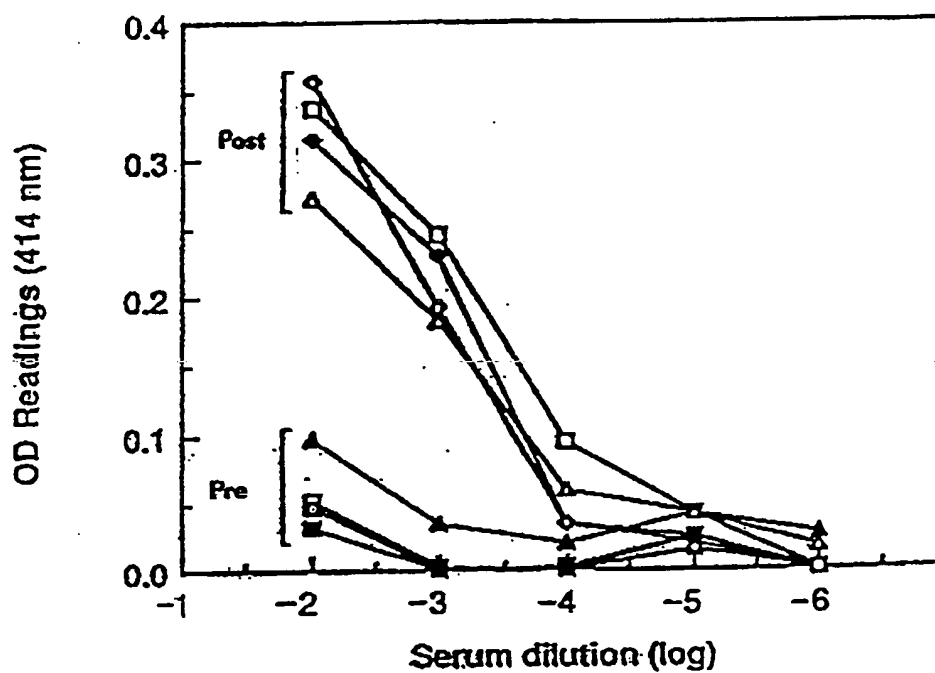
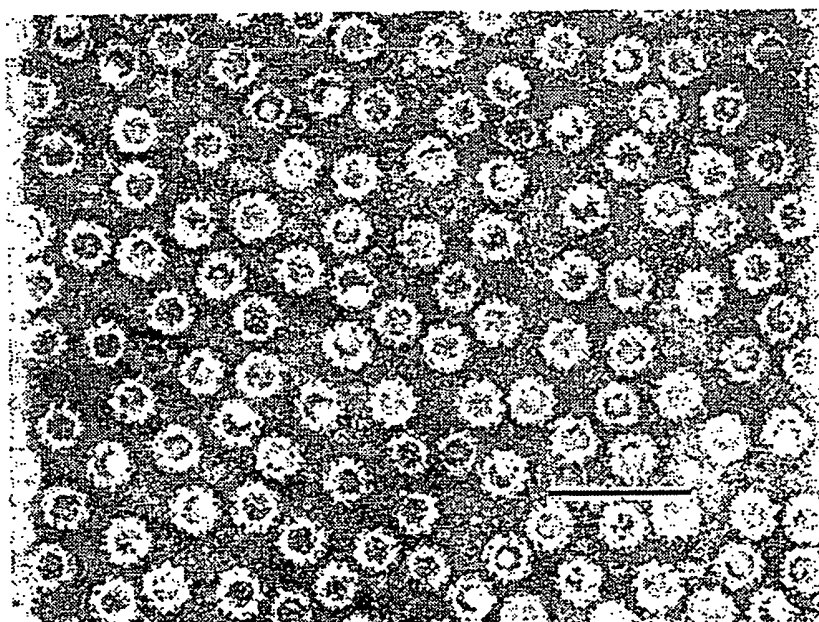


FIG 16

SUBSTITUTE SHEET

36/39

Figure 17



SUBSTITUTE SHEET

37/39

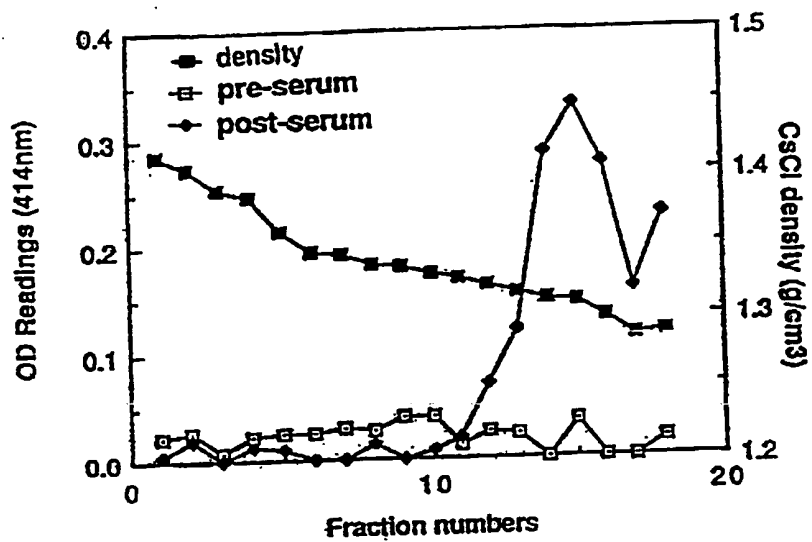
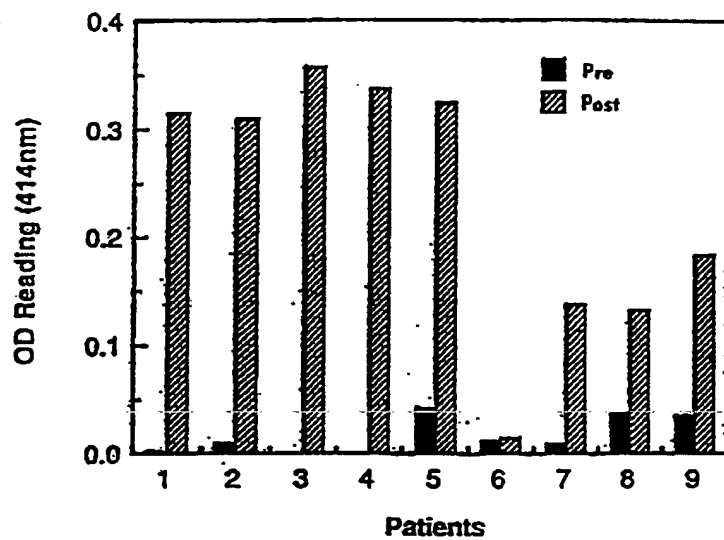


FIG 18

SUBSTITUTE SHEET

38/39

**FIG 19****SUBSTITUTEHEET**

TGACCGGACC TGCTGTGAA GATCTCTCA
AANGGCTCGA ACGACCAAG CACGATCGGT
ATTGTGTGA CTACGCAAG TGGGACTCAA
CCCANCCACC AAAAGTACA TCCAATCAAT
NGACATC....
GTGANATGNN ACATCTTGA CTCGATGGAC
CTATTCACAT ATGGTGATGA CGGTGCTAC
ATCGTCCAC CACTATATCA TCTGTCAATGC
CCAAGTCTTC ACCAACCTGA AAC

FIG 20